Published in Journals: Buildings, Sensors, Smart Cities, Sustainability, Urban Science and Applied Sciences

Topic Reprint

Smart Cities

Infrastructure, Innovation, Technology, Governance and Citizenship - Volume II

Edited by
Luis Hernández-Callejo, Sergio Nesmachnow and Pedro Moreno-Bernal

mdpi.com/topics

MDPI

Smart Cities: Infrastructure, Innovation, Technology, Governance and Citizenship—Volume II

Smart Cities: Infrastructure, Innovation, Technology, Governance and Citizenship—Volume II

Topic Editors

Luis Hernández-Callejo
Sergio Nesmachnow
Pedro Moreno-Bernal

MDPI

Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

Topic Editors
Luis Hernández-Callejo
Department of Agricultural and Forestry Engineering
University of Valladolid
Soria
Spain

Sergio Nesmachnow
Instituto de Computación
Universidad de la República
Montevideo
Uruguay

Pedro Moreno-Bernal
Faculty of Accounting, Administration, and Informatics, Autonomous University of Morelos State (UAEM)
Cuernavaca
Mexico

Editorial Office
MDPI AG
Grosspeteranlage 5
4052 Basel, Switzerland

This is a reprint of the Topic, published open access by the journals *Buildings* (ISSN 2075-5309), *Sensors* (ISSN 1424-8220), *Smart Cities* (ISSN 2624-6511), *Sustainability* (ISSN 2071-1050), *Urban Science* (ISSN 2413-8851) and *Applied Sciences* (ISSN 2076-3417), freely accessible at: https://www.mdpi.com/topics/1QQN8GMG0O.

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

Lastname, Firstname, Firstname Lastname, and Firstname Lastname. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

Volume II
ISBN 978-3-7258-5015-0 (Hbk)
ISBN 978-3-7258-5016-7 (PDF)
https://doi.org/10.3390/books978-3-7258-5016-7

Set
ISBN 978-3-7258-4951-2 (Hbk)
ISBN 978-3-7258-4952-9 (PDF)

© 2025 by the authors. Articles in this book are Open Access and distributed under the Creative Commons Attribution (CC BY) license. The book as a whole is distributed by MDPI under the terms and conditions of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) license (https://creativecommons.org/licenses/by-nc-nd/4.0/).

Contents

About the Editors . . . . . **vii**

Preface . . . . . **ix**

George N. Papageorgiou, Demetris Demetriou, Elena Tsappi and Athanasios Maimaris
Analyzing the Requirements for Smart Pedestrian Applications: Findings from Nicosia, Cyprus
Reprinted from: *Smart Cities* **2024**, *7*, 1950–1970, https://doi.org/10.3390/smartcities7040077 . . **1**

Haiyang Lyu, Qiqi Zhong, Donglai Jiao and Jianchun Hua
Bump Feature Detection Based on Spectrum Modeling of Discrete-Sampled, Non-Homogeneous Multi-Sensor Stream Data
Reprinted from: *Appl. Sci.* **2024**, *14*, 6744, https://doi.org/10.3390/app14156744 . . . . . **22**

Song Yang, Yingfeng Su and Qin Yu
Smart-City Policy in China: Opportunities for Innovation and Challenges to Sustainable Development
Reprinted from: *Sustainability* **2024**, *16*, 6884, https://doi.org/10.3390/su16166884 . . . . . **53**

El-hacen Diallo, Rouwaida Abdallah, Mohammad Dib and Omar Dib
Decentralized Incident Reporting: Mobilizing Urban Communities with Blockchain
Reprinted from: *Smart Cities* **2024**, *7*, 2283–2317, https://doi.org/10.3390/smartcities7040090 . . **85**

Ornella Salimbene, María Teresa Baeza-Romero, Francesco Pilla and Gregor Čok
Air Quality Awareness—Empirical Evidence from a Comparative Perspective between Two European Cities
Reprinted from: *Urban Sci.* **2024**, *8*, 133, https://doi.org/10.3390/urbansci8030133 . . . . . **120**

Flann Chambers, Giovanna Di Marzo Serugendo and Christophe Cruz
Autonomous Generation of a Public Transportation Network by an Agent-Based Model: Mutual Enrichment with Knowledge Graphs for Sustainable Urban Mobility
Reprinted from: *Sustainability* **2024**, *16*, 8907, https://doi.org/10.3390/su16208907 . . . . . **136**

Junhong Hu, Mingshu Yang, Yunzhu Zhen and Wenling Fu
Node Importance Evaluation of Urban Rail Transit Based on Signaling System Failure: A Case Study of the Nanjing Metro
Reprinted from: *Appl. Sci.* **2024**, *14*, 9600, https://doi.org/10.3390/app14209600 . . . . . **158**

Guobin Gu, Xin Sun, Benxiao Lou, Xiang Wang, Bingheng Yang, Jianqiu Chen, et al.
A Study of Mixed Non-Motorized Traffic Flow Characteristics and Capacity Based on Multi-Source Video Data
Reprinted from: *Sensors* **2024**, *24*, 7045, https://doi.org/10.3390/s24217045 . . . . . **175**

Zhipeng Fan, Yayun Wu, Wei Liu, Ming Chen and Zeguo Qiu
LG-YOLOv8: A Lightweight Safety Helmet Detection Algorithm Combined with Feature Enhancement
Reprinted from: *Appl. Sci.* **2024**, *14*, 10141, https://doi.org/10.3390/app142210141 . . . . . **198**

Ernesto De la Cruz-Nicolás, Hugo Estrada-Esquivel, Alicia Martínez-Rebollar, Odette Alejandra Pliego-Martínez and Eddie Clemente
Index for Assessing the Performance Level of Vehicular Traffic on Urban Streets
Reprinted from: *Urban Sci.* **2024**, *8*, 204, https://doi.org/10.3390/urbansci8040204 . . . . . **214**

Lapyote Prasittisopin
How 3D Printing Technology Makes Cities Smarter: A Review, Thematic Analysis, and Perspectives
Reprinted from: *Smart Cities* **2024**, *7*, 3458–3488, https://doi.org/10.3390/smartcities7060135 . . **236**

Sanjana Paul, Mikita Klimenka, Fabio Duarte, Carmen Crawford, Claire Gorman, Carlo Ratti and Jacopo Buongiorno
When Cities Go Nuclear: Exploring the Applications of Nuclear Batteries Toward Energy Transformation
Reprinted from: *Urban Sci.* **2024**, *8*, 226, https://doi.org/10.3390/urbansci8040226 . . . . . **267**

Diego Rossit, Begoña González Landín, Mariano Frutos and Máximo Méndez
Scatter Search Algorithm for a Waste Collection Problem in an Argentine Case Study
Reprinted from: *Urban Sci.* **2024**, *8*, 240, https://doi.org/10.3390/urbansci8040240 . . . . . **281**

Shabnam Varzeshi, John Fien and Leila Irajifar
Integrating Technology and Urban Resilience: A Comprehensive Analysis of Smart City Initiatives in Sydney
Reprinted from: *Sustainability* **2024**, *16*, 10967, https://doi.org/10.3390/su162410967 . . . . . **310**

Ouiam Fatiha Boukharta, Loïc Sauvée, Leticia Chico-Santamarta, Fabiana Pena-Fabri and Luis Manuel Navas-Gracia
Reality vs. Expectations in the Implementation of Urban Agricultural Projects—A Polycentric Governance Analysis
Reprinted from: *Urban Sci.* **2024**, *8*, 260, https://doi.org/10.3390/urbansci8040260 . . . . . **332**

Jing Xiong, Youchao Sun, Zhihao Xu, Yongbing Wan and Gang Yu
Remaining Life Prediction of Automatic Fare Collection Systems from the Perspective of Sustainable Development: A Sparse and Weak Feature Fault Data-Based Approach
Reprinted from: *Sustainability* **2025**, *17*, 230, https://doi.org/10.3390/su17010230 . . . . . **349**

Salit Azoulay Kochavi, Oz Kira and Erez Gal
Real-Time Monitoring of Environmental Parameters in Schools to Improve Indoor Resilience Under Extreme Events
Reprinted from: *Smart Cities* **2025**, *8*, 7, https://doi.org/10.3390/smartcities8010007 . . . . . **377**

Benjamin Neale, Sakdirat Kaewunruen, Dan Li, Ayfer Donmez Cavdar, Rand Askar, Tatjana Tambovceva and Diana Bajare
Challenges of Engineering Skillsets Essential for Driving Circularity of Smart Cities
Reprinted from: *Appl. Sci.* **2025**, *15*, 809, https://doi.org/10.3390/app15020809 . . . . . **395**

About the Editors

Luis Hernández-Callejo

Luis Hernández-Callejo is a full professor and researcher at Universidad de Valladolid, Spain. He has participated in numerous research projects and successfully supervised more than 10 doctoral theses. He is the editor-in-chief of several prestigious journals and a guest editor for numerous Special Issues in several other journals. His research interests include renewable energy, smart cities, microgrids, and artificial intelligence applied to engineering.

Sergio Nesmachnow

Sergio Nesmachnow is a full professor and researcher at Universidad de la República, Uruguay. He has more than 450 published articles across journals and international conferences. His main research areas are high performance computing, optimization, and smart cities.

Pedro Moreno-Bernal

Pedro Moreno-Bernal is a full-time Associate Professor at the Universidad Autónoma del Estado de Morelos and a National Research Associate at the Secretaría de Ciencia, Humanidades, Tecnología e Innovación in Mexico. His areas of interest include computational intelligence, metaheuristics, energy sustainability, and smart cities. He has co-authored multiple indexed scientific publications, 11 of which were indexed in Journal Citation Reports and 31 in Scopus, resulting in an h-index of 5. Furthermore, he has served as a guest editor for Special Issues in various journals.

Preface

Cities are constantly changing, and this is the main reason we are launching this Reprint of articles. The aim is to collect scientific articles on different topics that affect cities, and these topics mainly include energy, IoT, computing, and others.

The main purpose of this Reprint is to present some research advances across the above-mentioned topics, which can serve as a solid foundation for researchers who are starting to work on topics related to smart cities.

Luis Hernández-Callejo, Sergio Nesmachnow, and Pedro Moreno-Bernal
Topic Editors

smart cities

MDPI

Article

Analyzing the Requirements for Smart Pedestrian Applications: Findings from Nicosia, Cyprus

George N. Papageorgiou, Demetris Demetriou, Elena Tsappi and Athanasios Maimaris *

SYSTEMA Research Center, European University Cyprus, Nicosia 2404, Cyprus;
g.papageorgiou@euc.ac.cy (G.N.P.); demetriou.deme@gmail.com (D.D.); tsappielena@gmail.com (E.T.)
* Correspondence: a.maimaris@cycollege.ac.cy; Tel.: +357-22559488

Abstract: This paper elicits and analyzes the main requirements for Smart Pedestrian applications designed to enhance the pedestrian experience in urban environments by offering optimized walking routes, improved accessibility, and support for social inclusion and connectivity. Utilizing a mixed-methods approach, the research combines qualitative insights with quantitative data analysis based on surveys conducted in two strategically selected urban areas of Nicosia, Cyprus. Through the survey, the requirements and potential use of Smart Pedestrian apps are investigated while accounting for the quality of service of the urban infrastructure in a medium-sized city context. Additionally, the study contrasts the current smartphone applications, as they predominantly facilitate vehicular transportation, with the potential use of ICT/ITS to support pedestrians for sustainable mobility. The findings reveal a significant demand for a Pedestrian Smartphone app, driven by its ability to provide relevant information on optimum pedestrian routes, as well as act as a citizen's voice for spotting infrastructure problems and improving the pedestrian network. Further, it is also revealed that limitations in the pedestrian infrastructure substantially restrict walking preferences, emphasizing the need for urgent city-level urban planning solutions to support active mobility. Additionally, the research carried out underscores the importance of a sustainable business model to support the successful deployment of Smart Pedestrian apps. Ultimately, the results of the study suggest prioritizing a smart technology leverage with a crowdsourcing social network business model to promote pedestrian mobility, thereby reducing vehicular dependence, enhancing public health, and improving the quality of life. Such an approach would act as catalyst for policymakers to concentrate on sustainability by investing in digital technology for integrated pedestrian networks, fostering the emergence of genuine smart cities.

Keywords: sustainable mobility; smartphone apps development; pedestrian network; urban development; market research; information and communication technology (ICT); intelligent transportation systems (ITS); smart cities

1. Introduction

Across many urban areas in Europe and particularly in Cyprus, there is an extremely high overreliance on motorized fossil-fuel vehicles to carry out daily mobility activities. As a result of this dependency we have multiple negative socioeconomic, health and environmental impacts. Despite the recent rapid development of the smart city concept accompanied with advancements in Information and Communication Technology (ICT) and Intelligent Transportation Systems (ITS), the issue remains unresolved and, in many urban spaces, where smart technologies are employed is even worse. Typically, ICT and ITS applications are focused on promoting motorized traffic, often neglecting non-motorized, sustainable modes of transport such as walking and cycling [1–7]. Moreover, as urban areas evolve towards becoming "smarter" through ICT and ITS applications, temporary traffic improvements consequently lead to increased vehicle use, exacerbating congestion and creating a vicious circle where solutions further entrench the problem!

Additionally, the lack of adequate infrastructure, particularly in many medium-sized cities contributes to this issue. In Cyprus cities, for instance, the public transportation system is plagued by irregular and unreliable schedules with limited coverage, pushing residents towards personal vehicle ownership and use. This motorized vehicle reliance has positioned Cyprus as the fourth highest among thirty-four European countries in terms of car ownership, with 658 cars per 1000 residents in 2022, posing significant challenges such as public health deterioration, increased pollution, exacerbation of climate change, and threats to wildlife ecosystems [2].

Paradoxically, Cyprus's climatic conditions naturally support active transport modes such as walking and cycling. Therefore, an intriguing research question is what the factors are impeding the adoption of an active sustainable mobility lifestyle. Can technology support such sustainable mobility modes? Specifically, what are the requirements and features for a successful Pedestrian Smartphone app? Could the strategic use of ICT and ITS, especially through smartphone apps, promote a shift towards more active mobility? Identifying drivers that could foster such a shift is absolutely necessary, especially in naturally walkable cities like Nicosia, where the terrain is generally flat and distances are manageable. Further, it is noteworthy that such smartphone applications could contribute to a better sustainable touristic experience for the many visitors of Cyprus. As a result, active mobility could be promoted through smart solutions.

Transitioning to active mobility could offer manifold benefits for individuals, society, the economy, and the environment. It is important to note that the extensive use of motorized vehicles for daily transportation needs negatively impacts human health. Promoting active transportation would enhance physical activity and foster healthier lifestyles, moving away from the pervasive inactivity during daily commutes and other routine activities, which pose numerous health hazards. In today's digital society, a great opportunity emerges for employing appropriate ICT/ITS smartphone apps to facilitate active mobility. Such mobile apps could also be used for data collection and social marketing campaigns to positively change social norms and attitudes towards active transportation, which could be a pivotal move. These smart solutions for promoting active transportation would yield substantial benefits for individuals and the broader urban community, supporting a paradigm shift in transportation systems towards sustainable development in contemporary smart cities.

Specifically, smartphone navigation apps dedicated to walkability and sustainable mobility could act as a catalyst for changing mindsets [1]. A prior work focusing on evaluating user attitudes towards the quality of service in pedestrian networks revealed a significant demand for enhanced pedestrian infrastructure and identified key user requirements for a smart mobility solution [2]. However, it also highlighted the need for a more comprehensive analysis incorporating diverse urban settings and advanced data analysis, which the current study aims to address. Simultaneously, real-time information about the urban environment or the current climate crisis could further motivate this shift. Moreover, advancements in new technologies such as artificial intelligence (AI) and the big data analytics would enable governments and municipal institutions to efficiently collect vast amounts of data via crowdsourcing to properly assess pedestrian infrastructure conditions and walkability levels and effectively plan for sustainable mobility.

Despite extensive research into smart urban mobility, a significant gap still remains in revealing the needs for active transportation and developing solutions specifically for pedestrians and cyclists. As pointed out earlier, ICT and ITS solutions are mainly directed to satisfy the needs of motorized vehicle drivers rather than pedestrians. Further, despite some research projects and applications carried out in big cities around the world, there is limited practical, application-oriented research tailored to specific urban context settings and medium-sized cities around Europe. Prior studies [3,8,9] exploring pedestrian mobility and pedestrian network concepts also revealed this gap regarding the context of practical deployment, but yet, not much has been done.

This paper aims to bridge the identified gaps by carrying out market-oriented research investigating the requirements and the factors that would motivate active mobility and the role of a generic smartphone walkability application in this context. Taking as a case study a medium-sized walkable European city, data are collected via a survey questionnaire administered to the residents of Nicosia, Cyprus, providing a tangible assessment of the problem situation. Additionally, the survey explores the economic viability of pedestrian mobile apps in a competitive market where similar applications are often available for free. This approach fills a crucial empirical void in smart city research by focusing on pedestrian mobility and delivers detailed insights into both user acceptance and the economic implications of applying ICT and ITS technology solutions in this area.

The paper goes first through a review previous studies on the efficacy of technology and smartphone apps for promoting walkability and sustainable urban development. This is followed by a description of the survey methodology, including participant recruitment and challenges encountered. Then, the analysis of the collected data is carried out, presenting the main findings. Finally, the paper concludes with a discussion of the potential impact and success of an envisioned Smart Pedestrian application, synthesizing findings and implications for future smart urban mobility.

2. Review of Relevant Studies

In this section, first a critical examination of the scholarly discourse is carried out on the multidimensional benefits and barriers to pedestrian mobility. The objective is to identify the key preconditions that encourage pedestrian activities and to explore the challenges that may discourage individuals from engaging in or enjoying walkability. Following this, the review shifts focus to the methodologies behind the development of ICT/ITS smartphone applications, highlighting the crucial elements necessary for their successful implementation. The discussion culminates in a comprehensive analysis of successful smartphone applications, providing a detailed breakdown of the dynamics that contribute to their effectiveness. This examination offers insights into how digital tools facilitate pedestrian navigation and engagement, demonstrating their practical applications and theoretical implications.

2.1. Walkability Studies

The impact of a walkable urban space design on pedestrian behavior has been extensively studied, revealing significant benefits. Particularly, the research by [10,11] utilized a comprehensive walkability index incorporating land use diversity, street network design, residential density, and retail floor area ratios to assess health outcomes. The findings demonstrated that well-designed neighborhoods not only reduce the incidence of chronic conditions such as obesity but also contribute to lowering environmental pollution. Further, research has underscored walking as a vigorous exercise beneficial for cardiovascular health and helpful in mitigating diabetes and mental health disorders [9,11]. These activities are supported by evidence of the extensive health benefits provided by regular walking [11,12]. Additional studies have focused on aerobic fitness, particularly the enhancement of VO2max, highlighting its importance for efficient oxygen utilization during physical exertion [9].

In addition, the insights from [10] showed how Geographic Information Systems (GIS) can be utilized to assess urban environments to improve the human experience and facilitate active mobility. Key urban design elements like land use, accessibility, and activity variety are crucial in influencing walking decisions. These elements encourage short walking activities, especially in areas with dense intersections and a variety of shops and amenities. The integration of smart technologies, combined with effective governance and thoughtful platform design, would play a critical role in enhancing urban walkability, thus supporting broader health and environmental benefits [13]. Additionally, big data analytics have the potential to refine pedestrian pathways by optimizing safety and space

efficiency, incorporating information on essential elements such as pavements and street furniture, which are crucial for user-friendly and accessible pedestrian networks [4,5,14].

A diverse mix of land uses and visually appealing urban designs has been shown to create environments more conducive to walking. The availability of various walking routes to a destination also increases the likelihood of choosing to walk, underscoring the complex relationship between urban design and walking behavior. Thoughtfully designed urban spaces significantly encourage physical activity and, by extension, enhance public health. Moreover, economic studies have linked walkability to increased residential property values, depicting pedestrian-friendly infrastructures as both a public health enhancement and a strategic economic investment [15]. This highlights the multifaceted importance of walkability in improving quality of life and well-being, affirming the value of pedestrian-centric urban design in public health and city planning frameworks.

Further, the research led by Owen et al. [16] investigated the significant roles played by residents' socioeconomic backgrounds and specific neighborhood characteristics in four varied settings. This study found that income is a critical factor influencing walkability, more so than family status, which showed no significant correlation. Education was also seen as a positive influence, with more educated individuals being more aware of the personal and communal benefits of walking. Moreover, the study by Khajehpour and Miremadi [17] on equity challenges within bicycle-sharing systems provided valuable insights into procedural and recognition injustices, underscoring the importance of inclusive and equitable design in pedestrian network applications.

In another related study, Saelens et al. delved into the Neighborhood Environment Walkability Scale (NEWS), assessing its validity. The findings highlight how walkability is influenced by factors including the diversity of destinations, residential density, the quality of walking infrastructure, aesthetic considerations, traffic safety, and crime rates. In neighborhoods with lower walking levels, the aesthetic appeal was prioritized, whereas, in highly walkable areas, elements such as residential density and land use diversity were more favorably rated [18]. Finally, Ewing et al. identified key urban design attributes like imageability, visual enclosure, human scale, transparency, and complexity as significant in motivating walking. When effectively integrated into the urban fabric, these elements enhance the appeal and practicality of walking as a mode of transport [19].

This comprehensive review emphasizes that the decision to walk is influenced by a blend of urban design, personal health considerations, and quality of life factors. As urban landscapes continue to evolve, the need for designs that accommodate and encourage pedestrian activities remains critical [20]. The potential of utilizing ICT/ITS technology for managing the variety of influencing factors could act as a catalyst for enhancing walkability. This will be explored in more detail in the subsequent section.

2.2. Efficacy of ICT/ITS Mobile Applications

The efficacy of smartphone applications that leverage ICT and ITS to enhance walkability is closely linked to a sound business model [6,7,21,22]. Various strategies and features significantly influence application demand. Studies such as those by Ghose and Han [23] elucidate how factors like application size, age, description, and the number of screenshots correlate with usage and acceptance. Importantly, there is often a divergence between anticipated and actual application usage, which underscores a critical challenge in application development: aligning real-world use with developer expectations.

Further, the integration of sustainable urban planning with mobile application functionality is crucial. Wang et al. [24] illustrated how the sustainable integration of transport and urban planning enhances the applicability of mobile technologies, supporting sustainable urban living. This finding stresses the importance of strategic planning that aligns mobile applications with broader urban development goals to ensure that technology supports and enhances sustainable urban environments.

Furthermore, in an experimental study on the utility of smartphone apps [25], researchers determined whether an app could enhance the performance of medical profes-

sionals trained in advanced life support by providing easier access to critical information. Although the app was beneficial, it failed to become an indispensable resource, falling short of the initial product expectations. This highlights the crucial need for mobile applications to be adaptable and responsive to external conditions, a theme further explored by He et al. [26] during the COVID-19 pandemic. This illustrates the necessity for flexibility in mobile applications, enabling rapid responses to changes in user behavior and environmental conditions.

The above insights indicate that successful smartphone application development requires comprehensive requirement analysis during the design phase to create an app that effectively meets user needs. Specifically, Ries III et al. [27] explored the technical aspects of application development, emphasizing the importance of identifying the essential requirements and the risks associated with an inadequate requirement analysis or flawed system design. These potential pitfalls can lead to severe consequences, highlighting the critical role of meticulous design and testing phases in app development.

User feedback has emerged as a vital component in the development process, providing developers and researchers with invaluable insights that can significantly enhance the product's market fit. For example, Krebs and Duncan [28] conducted an evaluation focusing on user feedback, which provided deep insights into user perceptions and app appraisal, emphasizing the importance of addressing technical features during the requirements gathering phase to ensure functionality meets user expectations.

The proliferation of smartphone applications across various sectors provides valuable insights into critical success factors for application adoption. For instance, educational applications often act as middleware facilitating access to digital content, underscoring the importance of user-friendly and functional design to ensure effective usage [29]. An effective advertising campaign is crucial for driving downloads and ensuring sustained use, with the Mobile Marketing Association [30] providing a comprehensive array of strategies for crafting such campaigns. Carefully assessing these strategies to ensure they align with the application's marketing goals is crucial for influencing ongoing engagement and decision-making processes surrounding its purchase.

The analysis above synthesizes findings from various studies to provide a comprehensive view of the critical factors influencing the efficacy of ICT and ITS applications in promoting urban walkability. To further enhance these applications, developers should employ adaptive design principles that respond to user feedback and changing environmental conditions. Additionally, integrating applications with broader urban planning initiatives is crucial to ensure they contribute positively to sustainable urban development. It is also essential to prioritize inclusive design to address the diverse needs of urban populations, thereby avoiding procedural and recognition injustices. These strategies will ensure that ICT and ITS applications not only meet the current urban mobility challenges but are also poised to adapt to future developments in urban environments.

2.3. Smartphone Applications Focusing on Pedestrians

In recent years, there has been an increasing interest in walkability studies and developing smartphone apps to enhance the overall pedestrian experience [31–35]. Generally, applications designed for pedestrians aim to motivate individuals to increase walking for daily activities, as well as for physical exercise, and contribute to a higher well-being and quality of life.

Asimakopoulos et al. [36] applied the self-determination theory to explore motivation criteria, aiming to enhance user engagement for fitness tracking and health applications. Specifically, an on-diary study was conducted with 34 experienced users of Fitbit and Jawbone fitness trackers, who provided diary logs twice weekly over four weeks. These logs measured motivation and self-efficacy over time to establish how user experiences or attitudes towards healthcare technology impacted motivation. Their analysis shows that user motivation and self-efficacy are highly dependent on successful data integration, gamification, content design, context sensing, and appropriate motivational feedback.

However, the study provided limited evidence to support the correlation between specific features of fitness trackers and long-term user motivation.

In another study [37], the efficacy of smartphone applications in enhancing physical activity among male adults was assessed. Three hypotheses were tested:

H1. Receiving social feedback leads to higher step counts than no feedback;

H2. Social feedback results in higher step counts compared to solo feedback;

H3. Solo feedback leads to higher step counts than no feedback.

The findings suggest that, while smartphone applications that provide step counts can increase physical activity in young to middle-aged men, the addition of social feedback does not significantly impact results. This may be particularly relevant for inactive males with low levels of physical activity.

Further, Direito et al. [38] examined how smartphone applications could enhance physical fitness and activity among the youth. The app for IMproving FITness (AIMFIT) was used to evaluate the effectiveness of two popular off-the-shelf applications in improving cardiorespiratory fitness among young people. The findings suggest considerable potential for these applications to extend the reach of behavioral change interventions, although their standalone effectiveness remains unconfirmed. Despite the widespread availability of commercial smartphone applications, empirical evidence supporting their impact on targeted behaviors is scant.

Moreover, Mollee et al. [39] investigated technological features in smartphone applications that promote physical activity. The study aimed to understand the landscape of smartphone applications that encourage physical activity among healthy adults. They introduced a framework to determine the extent to which these applications incorporate various technological features. The study found that the most implemented features include user input, textual/numerical overviews of user behavior, sharing of achievements on social networks, and general physical activity advice. Less common were features such as adaptability, integration with external sources, gamification, punitive measures, or options to contact experts. The results suggest significant potential for enhancing application features to improve utilization rates given the capabilities of the current mobile technology. Additionally, Jee [40] reviewed two types of applications advancing physical activity to improve health in adults. The first type utilizes built-in smartphone sensors like GPS or accelerometers, while the second type relies on user-provided data such as activity duration, dates, and types. Bravata et al. [41] also examined applications that track steps via a pedometer to see if they motivate users to walk more.

It is noteworthy that such smartphone applications could contribute to a better sustainable touristic experience, which would be especially valuable for touristic destinations such as Cyprus with its many visitors. In recent years, numerous cities around the world have developed and implemented applications for touristic purposes that have significantly enhanced the visitor experience and active mobility. For instance, the "Paris City Guide App" in Paris provides detailed maps, itineraries, and offline access to crucial tourist information, resulting in high user satisfaction due to its comprehensive content and user-friendly interface [42]. Similarly, the "Rome Travel Guide" in Rome offers augmented reality features and historical information, which has been effective in enriching the cultural experience of visitors [43]. New York City's "NYC Go" app integrates event notifications, restaurant recommendations, and booking services, praised for its comprehensive integration of diverse services [44]. Tokyo's "Tokyo Handy Guide" stands out for its multilingual support and real-time public transportation updates, which are crucial for non-Japanese speaking tourists. Evaluations indicated a high appreciation for its ease of navigation and reliability [45]. These examples demonstrate how well-designed and evaluated touristic apps can significantly improve the overall visitor experience by addressing specific user

needs and integrating valuable features. As a result, active mobility is promoted through smart solutions.

Further, recent advancements in pedestrian navigation technologies highlight the importance of personalized and context-aware features. For example, apps that provide real-time environmental data, such as air quality and noise levels, alongside navigation assistance are found to be more appealing to users who prioritize health and sustainability [46,47]. These findings suggest that incorporating similar features in our app could significantly improve user engagement and adoption.

Moreover, the use of gamification elements and social interaction features has been shown to enhance user motivation and retention in travel and mobility apps [48]. Integrating such elements into our app could encourage more consistent use and promote a community-oriented approach to active mobility. Recent studies have also emphasized the potential of AI and wearable technologies to improve pedestrian navigation accuracy and user experiences. For instance, AI-enhanced inertial navigation systems can provide robust and accurate positioning even in complex urban environments [49]. These technologies can be leveraged to develop more sophisticated and user-friendly pedestrian navigation apps. By drawing on these insights from existing research, we can obtain ideas for the better design and implementation of a pedestrian navigation app that meets user needs and expectations, ultimately contributing to the promotion of active and sustainable urban mobility.

Progressive smart cities around the world, as demonstrated by Zeichner et al. [50], are leveraging smartphone applications to enhance pedestrian safety. Almost one-third of all road accidents in New York City are attributed to extensive mobile phone use. In response, the city has restructured selected roads and developed smartphone applications that alert users to potential dangers. This highlights the importance of understanding why citizens use mobile phones in specific contexts and how this knowledge can be used to further enhance pedestrian safety. In a similar study, Jacob et al. [51] developed a smartphone application to secure the traversing of pedestrians. Further, Ganti et al. [52] analyzed an application that addresses not only tangible road infrastructure elements but also intangible factors such as social urban conditions, emphasizing the complexity of the pedestrian experience through the digital customer journey.

Table 1 provides a comparative analysis of the main features found in pedestrian applications contrasted with potential features that can easily be implemented via ICT/ITS. Specifically, these potentials are contrasted with what three mainstream mobile app alternatives can offer: Google Maps, MapMyWalk, and Streetwise [53,54]. Each row in the table corresponds to a specific feature commonly found in pedestrian applications, while the columns represent the individual applications. The features assessed include route optimization, real-time updates, accessibility options, and community engagement functionalities. This comparison helps identify the strengths and limitations of each application, offering insights into their unique value propositions and identifying the main features that enhance their pedestrian experience.

The economic sustainability of any Smart Pedestrian app is a critical aspect of its development, particularly in a competitive market that includes free services such as Google Maps, MapMyWalk, and Streetwise. To differentiate a Smart Pedestrian applications and justify potential costs to users, many apps use a freemium business model. This model offers basic navigation services for free, while premium features—such as personalized route optimization, advanced safety alerts, and integration with public transportation—are available at a fee. Such a model leverages the unique capabilities of the app, distinguishing it from the general functionalities provided by existing applications.

Table 1. A comparative analysis of pedestrian application features.

Feature	ICT/ITS Potential Features	Google Maps	MapMyWalk	Streetwise
Real-time Navigation	✓	✓	✓	✓
Walking Routes	✓	✓	✓	✓
Pavement Quality Info	✓	✗ (Limited)	✗ (N/A)	✗ (N/A)
Estimated Travel Time	✓	✓	✓	✓
Safety Ratings	✓	✗ (Basic)	✗ (N/A)	✓
Weather Conditions	✓	✓	✗ (N/A)	✓
Public Amenities	✓	✓	✗ (N/A)	✗ (N/A)
Activity Tracking	✓	✗ (Limited)	✓	✗ (N/A)
User Feedback Feature	✓	✓	✓	✗ (N/A)
Accessibility	✓	✓	✗ (Limited)	✓
User-Friendly Interface	✓	✓	✓	✓

Moreover, successful app development should explore innovative revenue streams through partnerships with local businesses and municipal governments. Location-based advertising and promotions integrated into the app could drive user engagement while supporting local commerce. Collaborating with city planners to integrate the Smart Pedestrian application into broader smart city initiatives could provide a sustainable funding source, aligning the application's goals with public infrastructure projects.

These strategic approaches aim to position the application not only as a competitive entity in the market but also as a valuable tool for enhancing urban mobility and sustainability, thus offering a compelling alternative to free applications by adding distinct, value-added features that cater to specific user needs and community goals.

The above review enables us to consider the impact that technology has on individuals' decision to engage in physical activity and make use of pedestrian networks. It is clearly crucial when developing an application for walkability that accurate and relevant information is provided to citizens regarding mobility. To sum up, the following factors should be considered when developing Smart Pedestrian apps:

- Main App Features: Safety, Navigation, Convenience, Conviviality, Connectivity, and Data Collection (Crowdsourcing).
- Business model: The correct pricing of the application must be considered. The predefined factors have both positive and negative influences on the application adoption and usage.
- Demographics of the target market: The characteristics of the population that potentially will download, buy, and use the application.
- Correct evaluation of the urban environment: Evaluation of criteria for the application, based on the context of the specific city in which the app will be used.
- Pedestrian needs and adequate requirements the application should fulfil: A collection of data about existing applications, how people utilize them, what their expectations regarding smartphone applications are, and their views on how these applications will be beneficial to them.

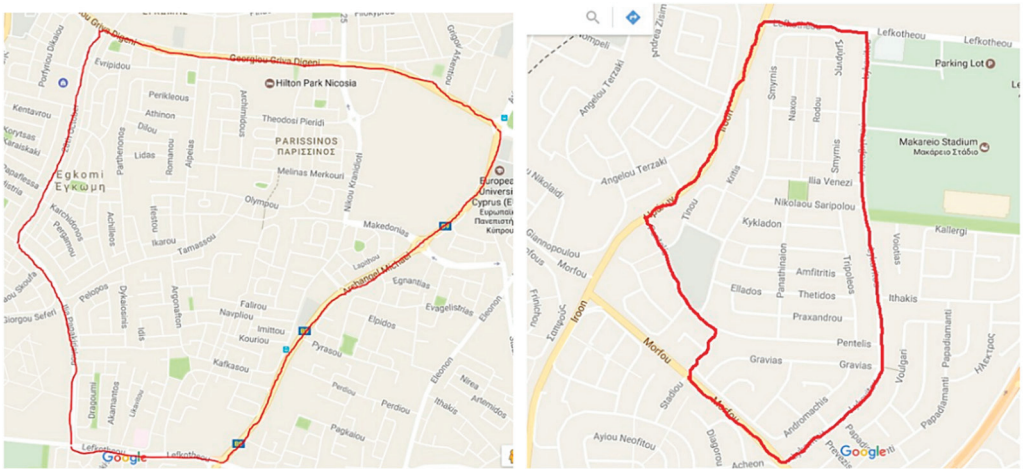
- Social marketing and communication techniques: Appropriate communication strategies must be considered. The predefined factors have both positive and negative influence on smartphone application adoption and usage.

The above factors should definitely be considered when deriving and analyzing the necessary requirements for successfully developing a Pedestrian Smart application in conjunction with the current issues faced by individuals (e.g., health), the natural environment (namely climate change and pollution), society, and the economy. The next section describes the methodology employed in an attempt to gather and analyze the complex spectrum of the necessary requirements for the case of the middle-sized city of Nicosia, Cyprus.

3. Methodology

In an effort to elicit and analyze the main requirements for a successful Smart Pedestrian app, a detailed on-site survey questionnaire was administered to residents of Nicosia, Cyprus. This survey utilized personal interview responses to a variety of questions (see Appendix A) to collect crucial data on walkability patterns, pedestrian infrastructure, amenities, and the potential use of a specialized Pedestrian Smartphone app.

The analysis carried out compares outcomes from two distinct urban areas of Nicosia—Area 1, characterized by its superior pedestrian infrastructure and accessible amenities, and Area 2, known for its inadequate conditions. Illustrated in Figure 1a,b, these areas were specifically chosen to explore the potential influence of differing qualities of pedestrian infrastructure on attitudes and preferences toward active mobility. Additionally, the study examined the possible benefits derived from employing ICT/ITS through the prospective Smart Pedestrian smartphone application, aiming to enhance these urban experiences.



(**a**) (**b**)

Figure 1. Selected urban areas within Nicosia covered by the survey [48]. (**a**) Area 1: High-Quality Service Area, (**b**) Area 2: Low-Quality Service Area.

The technical assessment of the Smart Pedestrian application's capabilities revealed its dynamic approach to enhancing urban walkability. Leveraging Geographic Information System (GIS) technology, the application dynamically maps and updates pedestrian pathways, employing sophisticated algorithms to analyze multiple variables—including path distance, surface conditions, and congestion—to recommend optimal walking routes. The integration of real-time data allows the application to adapt to temporary obstructions such as construction or events that may block pathways, incorporating updates from municipal sources and user feedback.

This dual approach to data collection—utilizing both municipal sources and user-generated content—ensures the delivery of accurate and timely information, thereby improving the overall user experience. The application's adaptability was further tested through simulated scenarios within the surveyed areas to assess its responsiveness and precision.

The methodology integrated into this study aimed to provide a comprehensive evaluation of pedestrian infrastructure effectiveness and potential user interactions with the application. The survey questionnaire was meticulously structured to collect data on demographics, walkability attitudes, and pedestrian environment assessments, along with inquiries about desired features and the usability of the mobile application. This multifaceted approach helped to achieve a deep understanding of the diverse factors influencing preferences for active mobility in modern urban environments.

Detailed contrasts between the two study areas highlighted the disparities in pedestrian network quality and simultaneously assessed the potential for an ICT/ITS Pedestrian Smartphone app. Area 1 featured well-maintained amenities, including nearby schools, universities, health centers, coffee shops, restaurants, and various public services that support pedestrian activities. In stark contrast, Area 2, with its limited services and amenities, underscored the challenges faced by pedestrians, particularly the elderly or disabled, due to its inferior infrastructure.

Subsequently, the results explored how variations in infrastructural and environmental conditions influence human behavior, social norms, and walking interactions. It aimed to determine whether such disparities affect human perception and, consequently, an individual's willingness to engage in walking. This information is invaluable for municipal institutions and policymakers, who can use the data to guide urban development and policy reforms based on public attitudes and infrastructural needs.

Further, the survey aimed to gauge potential user reactions to the proposed smartphone application, offering insights into its marketability and potential success in diverse urban conditions. The two areas served as a basis for developing a questionnaire designed to uncover current pedestrian needs and transportation preferences. After several revisions and a pilot study, the questionnaire was finalized into five major sections: demographics, walking mindset attitudes, pedestrian environment infrastructure, application features, and the desire for acquisition and use of the smartphone application. The demographic questions drew from previous surveys conducted by the Cypriot Statistical Service, while the other sections were informed by findings from prior literature reviews and current citizen needs. This comprehensive structure allowed for the collection of in-depth data on how potential users perceive existing infrastructural conditions and their receptiveness to technological solutions aimed at enhancing pedestrian mobility.

Utilizing a random sampling method, the survey was administered to participants in both Area 1 and Area 2. Of the 150 individuals approached, 108 (72%) provided responses—64 from Area 1 and 44 from Area 2—offering a comprehensive dataset for a detailed analysis. To ensure familiarity with the study areas, a preliminary screening verified each respondent's knowledge of both locations, enhancing the relevance and depth of the insights gathered. The response rate effectively illustrated the survey's resonance with the participants and underscored the pertinence of the questions posed. Employing a stratified random sampling technique, the study guaranteed a representative sample across key demographic characteristics such as age, gender, and socioeconomic status. It is important to note that the employed methodology represents a rigorous requirements elicitation approach, which is the primary step for any successful software development process.

4. Results: Data Analysis and Main Findings

The analysis of the collected data was conducted using IBM's Statistical Package for the Social Sciences (SPSS) V27. The investigation tested several hypotheses, including the correlation between pedestrian infrastructure quality, smartphone app use, and an individual's likelihood to engage in active mobility. The survey data from Area 2 showed that over 60% of pedestrians are significantly influenced by the conditions of the pedestrian network,

as well as meteorological factors. Similar findings were noted for Area 1, indicating a common tendency among individuals to consider weather conditions before opting to walk instead of using personal motorized vehicles. This is understandable, as under adverse weather conditions, walking becomes less appealing unless it is the only practical option.

The analysis confirmed the primary hypothesis (H1), suggesting a strong correlation between the quality of the pedestrian infrastructure and individuals' decision to walk. Additionally, the data revealed that the safety of walkable routes and the time of day significantly influence walking decisions. An inferential synthesis of these results led to insightful observations: the willingness to walk decreases as daylight fades, likely due to reduced perceptions of safety. Conversely, more daylight or earlier in the day tends to increase the appeal of walking. Interestingly, while factors like pavement quality and weather conditions do affect walking decisions, their impact is less pronounced than safety concerns. This outcome suggests that safety is the overriding factor, overshadowing any discomfort caused by adverse weather or poorly maintained pavements. Despite the challenges posed by inclement weather or suboptimal pavement conditions, these issues are minor compared to the overarching concerns about safety along pedestrian routes. This is particularly significant in Nicosia, where traffic dominated by motorized vehicles poses a great danger to pedestrians, who are not given much attention even when walking on zebra crossings.

The study also explored a second hypothesis regarding crowdsourcing and the dissemination of information about pedestrian conditions, adding depth to the comprehensive analysis of pedestrian mobility preferences. In Area 1, only 26% of respondents expressed willingness to purchase a smartphone pedestrian application priced between 0.01 and 0.99 EUR. The analysis showed a negative correlation between price and individuals' willingness to buy the application, indicating that, as the price increased, the likelihood of purchase decreased. A similar trend was observed in Area 2, where only 32% of participants were prepared to acquire the pedestrian mobile app within the same price range. These findings prompt several questions: Why are individuals deterred by such modest prices? Is it the presence of competing mobile applications available for free? Or do they have reservations about paying a monthly fee for mobile applications in general? Additionally, why does even a slight price increase lead to a significant drop in the willingness to purchase the application? While exploring these questions was beyond the scope of this paper, they merit further investigation to understand and potentially address these consumer behaviors.

Despite the above concerns, there was notable interest in purchasing the professional version of the proposed pedestrian application, with 45% in Area 1 and 46% in Area 2 indicating a willingness to purchase. This premium version of the app, which provides features of detailed information for connectivity, comfort, conviviality, and convenience, appealed to nearly half of the respondents in both areas, despite initial reservations about the pricing. Additionally, the freemium version of the smartphone app saw high potential user interest, with 70% of participants in Area 2 willing to download it, compared to 46% in Area 1. This suggests that individuals in areas with inferior infrastructure and fewer public services have a more acute need for the Smart Pedestrian application, relying more on the enhancements it offers compared to those in better-served areas. These insights underline a clear disparity in the perceived value and necessity of the application based on local infrastructure quality, and they reinforce the need for targeted strategies in application functionality that align with the specific needs and conditions of different urban environments. Also, the respondents of Area 2 seemed to look desperately for a solution to their mobility problems via the functionality of the Pedestrian Smartphone app. Regarding subscribing annually, which shows potential use, 14% of participants from Area 1 expressed a definite willingness to purchase the application compared to a slightly lower 11% from Area 2. This close similarity in the propensity to subscribe across both areas provides an intriguing insight, especially considering the higher level of interest from

individuals in Area 1. This appears to contradict the expectation that residents of Area 2, with its inferior infrastructure, would demonstrate a greater need for the subscription.

Further, the analysis assessed the impact of the proposed Smart Pedestrian smartphone application on walking behavior, aiming to understand how it might encourage walking either as a form of physical exercise or as part of daily routine activities. In Area 2 (despite the low-quality infrastructure), the data show that 50% of the respondents would use the Smart Pedestrian app for physical exercise, with 31.8% remaining undecided about their usage. Additionally, 46.5% reported they would use the Smart Pedestrian app for daily activities, though 23.3% were uncertain, and 30.7% indicated no interest in using the application. These results are encouraging for the potential successful deployment of a Smart Pedestrian application. However, they also raise important questions about why a significant portion of the sample displayed reluctance or uncertainty. Could their skepticism stem from negative experiences in poorly maintained areas, influencing their willingness to engage in outdoor exercise? Is it possible that the application does not fully meet their needs or expectations or that individuals prefer not to shift from their current exercise routines, such as gym workouts or other equipment-based activities? These considerations suggest that the Smart Pedestrian application could benefit from clearer communication of its features and benefits to better align with user needs.

In Area 1 (high-quality infrastructure), the results were slightly more positive, with 60% of respondents indicating they would use the application for daily walking, slightly higher than in Area 2. This suggests a more favorable reception to the application in areas with better infrastructure, highlighting the importance of aligning application features with the specific conditions and needs of different urban environments. The survey measurements indicate a divergence in attitudes towards the Smart Pedestrian application across the study areas. In Area 1, 25% of respondents are doubtful about using the Smart Pedestrian application, while 15% state definitively that they would not use it. Despite these reservations, the application is anticipated to be predominantly used for exercise, with a significant 60% of participants planning to use it for daily mobility-related activities. An additional 28.1% remain uncertain, highlighting an overall more positive attitude towards using the Smart Pedestrian application for physical exercise in Area 1. This inclination can be attributed to the high-quality infrastructure, which fosters a conducive environment for walking. It seems that features such as well-maintained pavements, accessible amenities, clear and easy-to-follow routes, and aesthetically pleasing surroundings encourage residents to engage in active mobility. Conversely, the absence of these ap-pealing features in other areas is a prohibiting factor, which also adds to fears of safety on the roads.

This survey study has also investigated how infrastructure conditions influence the possibility of adoption of the proposed Smart Pedestrian application in contrasting urban areas of low- and high-quality infrastructure. Interestingly, no significant correlation was found between regular walking habits and Smart Pedestrian app usage in either area. However, in Area 2, a significant positive correlation was observed between the willingness to use the application for both physical exercise and daily mobility needs at the 0.01 level. This suggests that the residents of areas of poorer infrastructure need a solution, probably through technology and crowdsourcing, to improve their walking experience.

Pedestrian App Use and Potential Interest in Social Networking Crowdsourcing

In terms of application use, residents from both areas show a general favorable inclination towards using the Smart Pedestrian application for physical exercise. However, the sentiment towards using the application for daily mobility activities differs markedly between the two areas, with Area 2 showing substantially less desire—2.5 times less compared to Area 1. This is expected, as the infrastructure conditions are comparatively adverse in Area 2. On the other hand, there is interest in using the app to voice concerns and problems of the pedestrian network via information crowdsourcing.

The analysis in Area 2 reveals a very strong correlation between the usage of the application for physical exercise and the activities of uploading and downloading information to and from the application. Similar strong correlations are noted for the application's use for daily walking, as evidenced by high Pearson correlation values that are statistically significant, with Sig (two-tailed) test values below the 0.01 level, indicating a 99% confidence in the validity of these results, as shown in Table 2.

Table 2. Low-quality service area correlation analysis.

		Crowdsourcing Uploading Info	**Regular Use Download Info**
Smart Pedestrian App Use for Physical Exercise	Pearson Correlation	0.682 **	0.719 **
	Sig (2-tailed)	0.000	0.000
	N	44	44
Smart Pedestrian App Use for Daily Mobility Needs	Pearson Correlation	0.611 **	0.596 **
	Sig (2-tailed)	0.000	0.000
	N	43	43

** Correlation is significant at the 0.01 level (2-tailed).

In Area 1 (see Table 3), as in Area 2, strong correlations were observed between the use of the Smart Pedestrian application for physical activities and the interactive behaviors of uploading and downloading information, although these correlations were slightly weaker in Area 1. The Pearson correlation values were statistically significant, with Sig (two-tailed) test values below 0.01, providing a 99% confidence level in the validity of these findings. These findings highlight the active engagement of residents in Area 1 with the application, demonstrating their propensity to both access and contribute information via crowdsourcing regarding their walking experience.

Table 3. High-quality service area correlation analysis.

		Crowdsourcing Uploading Info	**Regular Use Download Info**
Smart Pedestrian App Use for Physical Exercise	Pearson Correlation	0.496 **	0.441 **
	Sig (2-tailed)	0.000	0.000
	N	63	64
Smart Pedestrian App Use for Daily Mobility Needs	Pearson Correlation	0.492 **	0.519 **
	Sig (2-tailed)	0.000	0.000
	N	64	64

** Correlation is significant at the 0.01 level (2-tailed).

The strong correlation of app use to crowdsourcing likely arises from the expectation that local authorities will use the feedback to improve the pedestrian infrastructure. The slightly higher correlation coefficients found in Area 2 indicate a greater need for improvement in pedestrian conditions, prompting more frequent use of the application as a platform for expressing concerns and expectations. This suggests a proactive approach to community engagement, with residents actively adding to the data pool, thereby increasing the application's usefulness and relevance for all users. Such dynamic interaction underscores the application's potential as a crucial tool for the continuous development and enhancement of pedestrian networks.

Further analysis revealed that, in Area 2, there was a very strong correlation between using the application for physical exercise and the activities of uploading and downloading information. This was evidenced by high Pearson correlation values, which were statistically significant with Sig (two-tailed) test values below 0.01, indicating 99% confidence in

the validity of these results. Similar correlations were noted for the application's use for daily walking, further emphasizing the active involvement of Area 2 residents in utilizing the application for improving their walking experiences.

Summing up this Results section, it is important to note that the survey was instrumental in identifying and analyzing the main requirements for the proposed smartphone pedestrian application, as well as revealing its potential to significantly contribute to active transportation and, subsequently, to health and climate change mitigation. It was also shown that addressing infrastructural deficiencies and enhancing pedestrian safety are vital for encouraging walking and boosting the adoption of the Smart Pedestrian application. The analysis distinctively categorizes walking into two primary behaviors: walking for daily activity mobility needs and walking for physical exercise, each influenced by unique factors that affect the use of the Smart Pedestrian application. Daily mobility walking is characterized as a necessary and convenience-driven activity, often occurring in short, spontaneous bursts. This type of walking is highly dependent on the quality of urban infrastructure, as individuals are more likely to engage in walking when pathways and amenities are readily accessible and safe. Conversely, walking for physical exercise is a deliberate and planned activity, chosen primarily for its health benefits. This form of walking tends to occur at specific times and locations, often independent of direct environmental influences but still benefiting from high-quality pedestrian environments.

Recognizing these distinctions is critical for urban planners and application developers. It enables them to tailor environments and technological solutions that effectively support both types of walking activities. For instance, ensuring that daily paths are well-connected, safe, and integrated with essential services can enhance convenience-driven walking. Meanwhile, creating scenic and well-maintained routes can attract those walking for exercise. By differentiating these walking behaviors, urban planners and policymakers can better address the diverse needs of city residents regarding infrastructure, potentially increasing the adoption and effective use of apps like the Smart Pedestrian app, designed to improve urban pedestrian experiences. This approach not only caters to the immediate mobility needs but also promotes long-term health benefits and satisfaction with urban environments, fostering a more active and engaged society.

5. Discussion

This paper evaluated the main factors influencing walkability and the requirements for a successful Smart Pedestrian app. Via a survey study administered in Nicosia, Cyprus, various determinants of pedestrian behavior were assessed across two contrasting areas with distinctly different infrastructure qualities. The analysis effectively demonstrated the utility and appeal of the prospective Smart Pedestrian application, which aims to enhance urban walkability by offering deep insights into pedestrian behavior. Variations in the effect of the factors influencing walking decisions were detected, with connectivity, comfort, conviviality, conspicuity, and convenience being identified as pivotal. Further, the adoption of a freemium business model is strategically important, catering to users' expectations for free access to basic features while monetizing advanced capabilities. This model effectively addresses the prevalent resistance to paying for such applications.

Demand for the proposed smartphone app is prevalent; however, user engagement appears to vary in response to the quality of the pedestrian infrastructure, suggesting a complex market response. Further, approximately half of the respondents expressed a willingness to use the application for physical activity, highlighting a market segment receptive to the application's benefits for health and sustainable mobility. Also, safety was underscored as a critical concern influencing walking decisions, with a marked preference for well-lit streets and safer conditions.

The elicitation of the main requirements through the survey established the cornerstone for successfully developing the envisioned Smart Pedestrian application. It has also revealed its potential to significantly contribute to active transportation, health improvement, and climate change mitigation. The realization of this potential necessitates

a comprehensive understanding of the motivations driving walking within communities and the translation of these motivations into functional application features and the corresponding supportive urban planning strategies. Specifically, addressing infrastructural deficiencies and enhancing pedestrian safety are vital for encouraging walking and boosting the adoption of the proposed Smart Pedestrian application.

Note that the rigorous requirements engineering approach followed in this study bolsters the validity of the findings and improves the likelihood of developing a successful Pedestrian Smartphone application. Requirements elicitation is the first and most important step in the overall software development process of the application. This stage subsequently enables effective coding of the identified requirements through an agile prototyping method. For example, to optimize walking routes in urban environments, the route optimization features could easily be developed using Python and the NetworkX library, where pedestrian pathways can be analyzed and the shortest, safest, most convenient routes determined. Further, by utilizing established Geographic Information System (GIS) algorithm libraries, the envisioned application can dynamically map and update pedestrian pathways, employing sophisticated algorithms to analyze multiple variables—including path distance, surface conditions, and congestion—to recommend the optimal walking routes. Moreover, the integration of real-time data would allow the application to adapt to temporary obstructions such as construction or events that may block pathways, incorporating updates from municipal sources and user feedback. Such a dual approach to data collection—utilizing both municipal sources and user-generated content—would ensure the delivery of accurate and timely information, thereby improving the overall user experience. The application's adaptability could then be further tested through simulated scenarios within the surveyed areas to assess its responsiveness and precision.

Despite the significance of this study in the specific context of a medium-sized city, it is essential to recognize limitations in the generalizability of the results. This underscores the need for expanded research across various urban settings to strengthen the robustness and relevance of the findings. Future research could persist in examining pedestrian dynamics and integrating pedestrian-friendly policies to further clarify the intricate factors influencing urban walkability. Further, the survey inadvertently omitted areas that could provide additional insights, such as the reasons behind individuals' reluctance to pay for mobile apps. Understanding this reluctance is instrumental in developing appropriate business revenue models for the viability of any smartphone app. Additionally, a further demographic data analysis could have sharpened the findings, highlighting differences in application use and walking preferences across genders and age groups. This data could assist in identifying targeted audiences for the Smart Pedestrian application, facilitating a deeper understanding of varied walking behaviors and attitudes across demographic groups. Direct queries about desirable features in pedestrian applications would also provide developers with crucial information on user expectations and market gaps. Also, further research could establish the influence of human interactions, including whether family or friends using the Smart Pedestrian application would affect individual adoption decisions. On the other hand, this study advances the preliminary insights from previous research by providing a more granular analysis of pedestrian behavior across different urban settings [2]. The integration of the survey method and advanced data analysis in the software development process represents a significant methodological enhancement, offering a practical value for both application developers and urban planners. In this way, future research could continue to investigate the impact of advanced technologies such as augmented reality on pedestrian navigation and safety.

Despite the limitations, the study provides important insights into the current mindset and attitudes on pedestrian mobility and active transportation requirements. Further, it provides a robust foundation for further work and informs stakeholders about the potential market viability of a Smart Pedestrian application and the intricate role of municipality authorities. Interestingly, the analysis highlighted several findings: while individuals in areas with better infrastructure demonstrated a higher inclination to use the Smart

Pedestrian application, they are not the ones who most need it. Specifically, the individuals in Area 1, characterized by superior pedestrian amenities, are more willing to engage with the app due to more enjoyable experiences walking the streets compared to those in Area 2, where the pedestrian infrastructure is notably poorer. This reveals a fundamental difference within the findings. Although individuals in well-equipped areas like Area 1 seem more inclined to use the Smart Pedestrian application, it is actually the residents of less developed areas such as Area 2, who would benefit most from improved walkability, that need it the most. This discrepancy raises critical questions about the application's effectiveness in solving the issues it aims to address—primarily, to enhance pedestrian infrastructure and encourage walking in all urban areas, not just those already well serviced.

Additionally, the study establishes that infrastructural quality is a decisive factor influencing the Smart Pedestrian application's adoption and success. While the Smart Pedestrian app may promote walking in high-quality areas, its potential impact in low-quality zones remains limited if the demand is absent, thus not fully resolving the problem it was designed to tackle. This scenario points to a "half-done job", suggesting that the success and usage levels of the Smart Pedestrian app are also dependent on local infrastructure conditions. The identification of such problems underscores the importance of these factors for the Smart Pedestrian application's broader application and success. Before its deployment, it is essential for policymakers and governments to address infrastructural deficits to foster more inclusive walking encouragement across diverse urban areas. Additionally, the study brings to light privacy and surveillance concerns related to the Smart Pedestrian application. In an era, prone to technological breaches, ensuring the maximum protection of user privacy is critical. Without robust safeguards, potential users may be deterred from adopting the application, undermining its utility and reach.

Finally, it is important to note that the results from this study align with established findings in the realm of walkability and urban design, which highlight the critical influence of mixed land use, accessibility, safety, and connectivity regarding active mobility. However, this study advances the discourse by demonstrating significant context variations in the acceptance and effectiveness of Smart Pedestrian applications, suggesting that urban design principles cannot be uniformly applied but must be adapted to specific community characteristics and needs. Further, the presented nuanced analysis challenges the traditional perceptions of urban planners' role in promoting pedestrian mobility, emphasizing a strategy tailored to unique community characteristics over a one-size-fits-all solution. By doing so, the study contributes to the theoretical understanding of urban walkability and provides practical insights for the effective development and deployment of prospective Smart Pedestrian applications, ensuring it meets the diverse needs of people in contemporary urban spaces.

6. Conclusions

This paper highlights the critical role of innovative ICT/ITS and smart technology in fostering pedestrian-friendly environments and promoting active sustainable mobility. The insights gleaned from the potential use of ICT and ITS emphasized the benefits of urban connectivity and mixed-use developments, making a distinct contribution to the field by exploring the market potential for a Smart Pedestrian application in Nicosia, Cyprus. This study distinguishes itself from previous research by providing a rigorous methodology employing valid survey sampling procedures in requirements management, thus gaining actionable insights into the adoption of smart urban mobility solutions. Such a method is essential for a successful software engineering prototype-driven approach where real-world application and user feedback are integral to the development process.

Further, the deployment and economic considerations of the proposed application are addressed in a competitive market, where similar applications are freely available, by proposing a viable business model that underscores unique value propositions such as enhanced safety features, integration with public transport systems, and personalized route optimization. Thus, this research study fills a significant gap in empirical smart city

research but also offers a scalable model for other urban areas aiming to enhance pedestrian mobility through technology.

Looking ahead, potential developments for the proposed Smart Pedestrian application include integrating more sophisticated technologies and functionalities to enrich the user experience and promote urban walkability. Future development iterations might incorporate real-time data analytics to provide dynamic routing adjustments based on the current pedestrian traffic conditions, weather, and urban events. Exploring the use of augmented reality (AR) or mixed reality (MR) to offer immersive navigational cues and safety alerts could further enhance interactivity and user engagement. Moreover, expanding the collaboration with urban planners and local authorities to integrate infrastructural feedback mechanisms within the app could enable users to report issues directly to city management systems, significantly refining pedestrian pathways and addressing urban mobility challenges more effectively.

In conclusion, this study has demonstrated the tremendous potential of the proposed smart application to enhance urban walkability and has identified areas for further development and research. Particularly, for future research, we propose the integration of advanced analytical techniques and cross-disciplinary approaches to enrich the scientific contributions of our findings. Further work could also be done by adopting a longitudinal study design approach and expand the analysis to multiple cities for comparative analysis. Additionally, incorporating qualitative data through in-depth interviews and stakeholder focus groups, as well as utilizing emerging technologies such as the Internet of Things (IoT) for data collection, could offer deeper insights into user behavior and preferences. These enhancements would elevate the methodological rigor and practical relevance of the study, bridging the gap between traditional survey methods and innovative research practices in urban mobility. Such advancements will ensure that the findings contribute significantly to both academic discourse and practical urban planning strategies, promoting a more informed and successful implementation of pedestrian networks for the effective development of sustainable smart cities.

Author Contributions: The article is based on the MBA thesis of D.D. performed under the supervision of G.N.P. The questionnaire was conceived, distributed and analyzed by G.N.P., D.D. and E.T. A.M. wrote, reviewed, and edited the paper; visualization by A.M. and D.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data supporting the findings of this study are derived from a structured survey questionnaire administered to residents of two urban areas in Nicosia, Cyprus. The survey data, collected and analyzed as part of this research, involved responses that are not publicly available due to ethical restrictions and the privacy of the research participants. Data are available from the authors upon reasonable request and with permission of the European University Cyprus. Further details on the methodology and data collection can be found in the MDPI Research Data Policies section available at MDPI Research Data Policies.

Acknowledgments: The research presented in this paper is co-funded by the Republic of Cyprus and the European Regional Development Fund as part of ERA-NET Cofund Smart Urban Futures (ENSUF) Joint Programming Initiative (JPI) Urban Europe through the Research Promotion Foundation, protocol no. ΚΟΙΝΑ/ΓΙΚΠ URBAN EUROPE/1215/11. This framework is supported by the European Commission and funded under the HORIZON 2020 ERA-NET Cofund scheme. The funding covers the costs to publish in open access.

Conflicts of Interest: The authors declare no conflicts of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of the data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

Questionnaire

Demographics
Gender:
☐ Male ☐ Female
Age Group:
☐ $\leq$ 19 ☐ 20–24 ☐ 25–29 ☐ 30–34 ☐ 35–39 ☐ 40–44 ☐ 45–49 ☐ 50–54 ☐ 55–59 ☐ 60–64 ☐ $\geq$ 65
Nationality:
☐ Cypriot ☐ Other (please specify): _____
Highest Level of Education Completed:
☐ Primary ☐ Secondary ☐ Tertiary ☐ Post-graduate
Frequency of Walking
1i. As part of physical exercising: ☐ Never ☐ 1–3 times a week ☐ 4–6 times a week ☐ Daily
2ii. As part of fulfilling daily mobility needs (e.g., going to the supermarket, kiosk, coffee houses, restaurants, etc.): ☐ Never ☐ 1–3 times a week ☐ 4–6 times a week ☐ Daily
Opinion on walking's health benefits:
☐ Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree
Factors Influencing Walkability
2i. Time availability for walking: ☐ Very low ☐ Low ☐ Neutral ☐ High ☐ Very high
2ii. Pavement quality: ☐ Very low ☐ Low ☐ Neutral ☐ High ☐ Very high
2iii. Impact of weather conditions: ☐ Very low ☐ Low ☐ Neutral ☐ High ☐ Very high
7iv. Pedestrian Safety: ☐ Very low ☐ Low ☐ Neutral ☐ High ☐ Very high
Vehicle Use
I frequently use a motorized vehicle: ☐ Strongly Disagree ☐ Disagree ☐ Neutral ☐ Agree ☐ Strongly Agree
Smartphone Use
Do you own a Smartphone? ☐ Yes ☐ No
If yes, please answer the following:
How many apps have you installed on your smartphone? ☐ 0–4 ☐ 5–9 ☐ 10–14 ☐ 15–19 ☐ $\geq$20
How many apps do you use regularly?
☐ 0–4 ☐ 5–9 ☐ 10–14 ☐ 15–19 ☐ $\geq$20
Do you use any apps specifically for walking?
☐ Yes ☐ No

If yes, please answer the following:
3i. How many walking-related apps do you use? ☐ 1–3 ☐ 4–6 ☐ 7–9 ☐ $\geq$10
3ii. Frequency of use for these walking apps: ☐ Very low ☐ Very high 13iii. Please list the walking apps you use:
3iv. What types of information do these walking apps provide: ☐ Calories Burnt ☐ Distance Covered ☐ Number of Steps ☐ Other (Please Specify):
Continued interest in potential walking app features:
How many people are on the pavement: ☐ Broken Pavement ☐ Narrow/Wide Pavement ☐ Construction works ☐ Pavement Lighting
To what extent would you use it to:
4i. Update pavement information: ☐ Never ☐ Very Frequently
4ii. Be informed of pavement status: ☐ Never ☐ Very Frequently
Financial willingness for app features:
How much would you be willing to spend on such an app? 5i. €0.00–€0.99 USD: ☐ Never Pay ☐ Absolutely Sure 5ii. €1.00–€1.99 USD: ☐ Never Pay ☐ Absolutely Sure 15iii. €2.00–€2.99 USD: ☐ Never Pay ☐ Absolutely Sure 5iv. €3.00–€3.99 USD: ☐ Never Pay ☐ Absolutely Sure 5v. €4.00–€4.99 USD: ☐ Never Pay ☐ Absolutely Sure
To what extent are you willing to purchase a professional version of the application, which provides extra features such as calories burned, number of steps, and distance traveled? ☐ Never Pay ☐ Absolutely Sure
To what extent are you willing to pay for a yearly subscription to get more features? ☐ Never Pay ☐ Absolutely Sure
To what extent are you willing to download the application if it was free? ☐ Never ☐ Absolutely Sure
How much would the use of such an application affect your walking habits?
6i. For physical exercising: ☐ Never ☐ Absolutely Sure 6ii. For fulfilling daily activity mobility needs: ☐ Never ☐ Absolutely Sure

⊗ c 2024 by the authors. Submitted to *Future Internet* for possible open-access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/, (accessed on 9 January 2024).

References

1. Papageorgiou, G.; Maimaris, A. Towards the Development of Intelligent Pedestrian Mobility Systems (IPMS). In Proceedings of the 2017 International Conference on Electrical Engineering and Informatics (ICELTICs), Banda Aceh, Indonesia, 18–20 October 2017; pp. 251–256. [CrossRef]
2. Papageorgiou, G.; Maimaris, A.; Efstathiadou, T.; Balamou, E. Evaluating Attitudes on the Quality of Service of Pedestrian Networks. *WIT Trans. Built Environ.* **2017**, *176*, 35–41. [CrossRef]
3. Papageorgiou, G.; Demetriou, D.; Balamou, E.; Maimaris, A. Market Research and Concept Study for a Smart Pedestrian Network Application. In Proceedings of the 11th Pervasive Technologies Related to Assistive Environments Conference, Corfu, Greece, 26–29 June 2018; pp. 522–528.
4. Papageorgiou, G.; Petrakis, C.; Maimaris, A. Developing a smart pedestrian network big data platform for municipal organizations. In Proceedings of the 2nd European Conference on Electrical Engineering and Computer Science (EECS), Bern, Switzerland, 20–22 December 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 274–278.

5. Papageorgiou, G.; Hadjigeorgiou, K.; Ness, A.N. A big data approach to developing a smart pedestrian network (SPN) system. *WSEAS Trans. Environ. Dev.* **2019**, *15*, 371–378.
6. Conticelli, E.; Maimaris, A.; Papageorgiou, G.; Tondelli, S. Planning and designing walkable cities: A smart approach. In *Smart Planning: Sustainability and Mobility in the Age of Change*; Springer: Cham, Switzerland, 2018; pp. 251–269.
7. Papageorgiou, G.; Ioannou, A.; Maimaris, A.; Ness, A. A strategic approach for implementing a smart pedestrian network (SPN) system. *Open Comput. Sci.* **2021**, *11*, 208–217. [CrossRef]
8. Eurostat. Passenger Cars Per 1000 Inhabitants Reached 560 in 2022. Available online: https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20240117-1 (accessed on 9 January 2024).
9. Jabbari, M.; Fonseca, F.; Smith, G.; Conticelli, E.; Tondelli, S.; Ribeiro, P.; Ahmadi, Z.; Papageorgiou, G.; Ramos, R. The Pedestrian Network Concept: A Systematic Literature Review. *J. Urban Mobil.* **2023**, *3*, 100051. [CrossRef]
10. Frank, L.D.; Greenwald, M.J.; Kavage, S.; Devlin, A. *An Assessment of Urban Form and Pedestrian and Transit Improvements as an Integrated GHG Reduction Strategy*; Technical Report WA-RD 765.1; Washington (State) Department of Transportation Office of Research and Library Services: Washington, DC, USA, 2011; Submitted to Future Internet. Version 24 February 2024.
11. Morris, J.N.; Hardman, A.E. Walking to Health. *Sports Med.* **1997**, *23*, 306–332. [CrossRef] [PubMed]
12. American Hiking Society. Health Benefits of Hiking. Available online: https://americanhiking.org/resources/health-benefits-of-hiking/ (accessed on 7 March 2023).
13. Komninos, N.; Kakderi, C. (Eds.) *Smart Cities in the Post-Algorithmic Era: Integrating Technologies, Platforms, and Governance*; Edward Elgar Publishing: Cheltenham, UK, 2019.
14. Carter, E.; Adam, P.; Tsakis, D.; Shaw, S.; Watson, R.; Ryan, P. Enhancing Pedestrian Mobility in Smart Cities Using Big Data. *J. Manag. Anal.* **2020**, *7*, 173–188. [CrossRef]
15. Cortright, J. *Walking the Walk: How Walkability Raises Home Values in US Cities*; Technical Report; CEOs for Cities: Chicago, IL, USA, 2009.
16. Owen, N.; Cerin, E.; Leslie, E.; du Toit, L.; Coffee, N.; Frank, L.D.; Bauman, A.E.; Hugo, G.; Saelens, B.E.; Sallis, J.F. Neighborhood Walkability and the Walking Behavior of Australian Adults. *Am. J. Prev. Med.* **2007**, *33*, 387–395. [CrossRef]
17. Khajehpour, B.; Miremadi, I. Assessing Just Mobility Transitions in the Global South: The Case of Bicycle-Sharing in Iran. *Energy Res. Soc. Sci.* **2024**, *110*, 103435. [CrossRef]
18. Saelens, B.E.; Sallis, J.F.; Black, J.B.; Chen, D. Neighborhood-Based Differences in Physical Activity: An Environment Scale Evaluation. *Am. J. Public Health* **2003**, *93*, 1552–1558. [CrossRef]
19. Ewing, R.; Handy, S.; Brownson, R.C.; Clemente, O.; Winston, E. Identifying and Measuring Urban Design Qualities Related to Walkability. *J. Phys. Act. Health* **2006**, *3*, S223–S240. [CrossRef]
20. Miskolczi, M.; Földes, D.; Munkácsy, A.; Jászberényi, M. Urban Mobility Scenarios until the 2030s. *Sustain. Cities Soc.* **2021**, *72*, 103029. [CrossRef]
21. Papageorgiou, G.; Balamou, E.; Maimaris, A. Developing a business model for a smart pedestrian network application. In Proceedings of the Fourth International Congress on Information and Communication Technology: ICICT 2019, London, UK, 27–28 February 2019; Springer: Singapore, 2020; Volume 2, pp. 375–381.
22. Papageorgiou, G.; Petrakis, C.; Ioannou, N.; Zagarelou, D. Effective business planning for sustainable urban development: The case of active mobility. In Proceedings of the ECIE 2019 14th European Conference on Innovation and Entrepreneurship, Kalamata, Greece, 19–20 September 2019; Volume 2, p. 759.
23. Ghose, A.; Han, S.P. Estimating Demand for Mobile Applications in the New Economy. *Manag. Sci.* **2014**, *60*, 1470–1488. [CrossRef]
24. Wang, S.; Wang, S.; Wang, H. Transport-land use systems of sustainable London city. *Eco Cities* **2024**, *4*, 2379. [CrossRef]
25. Low, D.; Clark, N.; Soar, J.; Padkin, A.; Stoneham, A.; Perkins, G.D.; Nolan, J. A Randomised Control Trial to Determine If Use of the iResus© Application on a Smartphone Improves the Performance of an Advanced Life Support Provider in a Simulated Medical Emergency. *Anaesthesia* **2011**, *66*, 255–262. [CrossRef] [PubMed]
26. He, L.; Li, W.; Li, J.; Sun, J. Urban mobility analytics amid COVID-19 pandemic: A framework for promoting work resumption based on mobile phone data. *J. Transp. Geogr.* **2024**, *117*, 103887. [CrossRef]
27. Ries, K.G., III; Guthrie, J.G.; Rea, A.H.; Steeves, P.A.; Stewart, D.W. *StreamStats: A Water Resources Web Application*; Technical Report; US Geological Survey: Reston, VA, USA, 2008.
28. Krebs, P.; Duncan, D.T. Health App Use among US Mobile Phone Owners: A National Survey. *JMIR mHealth uHealth* **2015**, *3*, e101. [CrossRef] [PubMed]
29. Bomhold, C.R. Educational Use of Smartphone Technology: A Survey of Mobile Phone Application Use by Undergraduate University Students. *Program* **2013**, *47*, 424–436. [CrossRef]
30. Mobile Marketing Association. Available online: www.mmaglobal.com/files/mobileapplications.pdf (accessed on 7 April 2024).
31. Papageorgiou, G.; Demetriou, G. Examining the Impact of Learning Strategies on Developing Sustainable Mobility Mindsets: A Systems Approach. *WSEAS Trans. Environ. Dev.* **2019**, *15*, 533–543.
32. Fonseca, F.; Ribeiro, P.; Jabbari, M.; Petrova, E.; Papageorgiou, G.; Conticelli, E.; Tondelli, S.; Ramos, R. Smart Pedestrian Network: An integrated conceptual model for improving walkability. In Proceedings of the International Conference on Society with Future: Smart and Liveable Cities, Braga, Portugal, 4–6 December 2019; Springer International Publishing: Cham, Switzerland, 2019; pp. 125–142.

33. Fonseca, F.; Papageorgiou, G.; Tondelli, S.; Ribeiro, P.; Conticelli, E.; Jabbari, M.; Ramos, R. Perceived walkability and respective urban determinants: Insights from Bologna and Porto. *Sustainability* **2022**, *14*, 9089. [CrossRef]
34. Fonseca, F.; Ribeiro, P.J.; Conticelli, E.; Jabbari, M.; Papageorgiou, G.; Tondelli, S.; Ramos, R.A. Built environment attributes and their influence on walkability. *Int. J. Sustain. Transp.* **2022**, *16*, 660–679. [CrossRef]
35. Westenhöfer, J.; Nouri, E.; Reschke, M.L.; Seebach, F.; Buchcik, J. Walkability and urban built environments—A systematic review of health impact assessments (HIA). *BMC Public Health* **2023**, *23*, 518. [CrossRef]
36. Asimakopoulos, S.; Asimakopoulos, G.; Spillers, F. Motivation and User Engagement in Fitness Tracking: Heuristics for Mobile Healthcare Wearables. *Informatics* **2017**, *4*, 5. [CrossRef]
37. Harries, T.; Eslambolchilar, P.; Rettie, R.; Stride, C.; Walton, S.; van Woerden, H.C. Effectiveness of a Smartphone App in Increasing Physical Activity Amongst Male Adults: A Randomised Controlled Trial. *BMC Public Health* **2016**, *16*, 925. [CrossRef]
38. Direito, A.; Jiang, Y.; Whittaker, R.; Maddison, R. Apps for IMproving FITness and Increasing Physical Activity Among Young People: The AIMFIT Pragmatic Randomized Controlled Trial. *J. Med. Internet Res.* **2015**, *17*, e210. [CrossRef]
39. Mollee, J.S.; Middelweerd, A.; Kurvers, R.L.; Klein, M.C.A. What Technological Features Are Used in Smartphone Apps That Promote Physical Activity? A Review and Content Analysis. *Pers. Ubiquitous Comput.* **2017**, *21*, 633–643. [CrossRef]
40. Jee, H. Review of Researches on Smartphone Applications for Physical Activity Promotion in Healthy Adults. *J. Exerc. Rehabil.* **2017**, *13*, 3. [CrossRef] [PubMed]
41. Bravata, D.M.; Smith-Spangler, C.; Sundaram, V.; Gienger, A.L.; Lin, N.; Lewis, R.; Stave, C.D.; Olkin, I.; Sirard, J.R. Using Pedometers to Increase Physical Activity and Improve Health: A Systematic Review. *JAMA* **2007**, *298*, 2296–2304. [CrossRef]
42. Yang, S.; Hwang, W. *Designing Apps for Tourists: A Case Study*; Springer: Cham, Switzerland, 2014.
43. Walls, D.M. User Experience in Social Justice Contexts. In Proceedings of the 34th ACM International Conference on the Design of Communication, Silver Spring, MD, USA, 23–24 September 2016.
44. Hossain, M.J.; Islam, M.S. Selective Dissemination of Information (SDI) Service: A Conceptual Paradigm. *Int. J. Inf. Sci. Manag.* **2012**, *6*, 27–44.
45. Witten, I.H.; Loots, M.; Trujillo, M.F.; Bainbridge, D. The Promise of Digital Libraries in Developing Countries. *Commun. ACM* **2001**, *20*, 7–13. [CrossRef]
46. Gavalas, D.; Konstantopoulos, C.; Mastakas, K.; Pantziou, G. Mobile recommender systems in tourism. *J. Netw. Comput. Appl.* **2014**, *39*, 319–333. [CrossRef]
47. Fonseca, F.; Conticelli, E.; Papageorgiou, G.; Ribeiro, P.; Jabbari, M.; Tondelli, S.; Ramos, R. Use and Perceptions of Pedestrian Navigation Apps: Findings from Bologna and Porto. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 446. [CrossRef]
48. Deterding, S.; Dixon, D.; Khaled, R.; Nacke, L. From game design elements to gamefulness: Defining gamification. In Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, Tampere, Finland, 28–30 September 2011; pp. 9–15.
49. Prandi, C.; Barricelli, B.R.; Mirri, S.; Fogli, D. Accessible wayfinding and navigation: A systematic mapping study. *Univers. Access Inf. Soc.* **2023**, *22*, 185–212. [CrossRef]
50. Zeichner, N.; Perry, P.; Sita, M.; Barbera, L.; Nering, T. *Exploring How Mobile Technologies Impact Pedestrian Safety*; NYC Media Lab Research Brief: New York, NY, USA, 2014.
51. Jacob, R.; Mooney, P.; Corcoran, P.; Winstanley, A.C. Integrating Haptic Feedback to Pedestrian Navigation Applications. In Proceedings of the GISRUK 2011: GIS Research UK 19th Annual Conference, Portsmouth, UK, 27–29 April 2011; pp. 205–210.
52. Ganti, R.K.; Pham, N.; Ahmadi, H.; Nangia, S.; Abdelzaher, T.F. GreenGPS: A Participatory Sensing Fuel-Efficient Maps Application. In Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services, San Francisco, CA, USA, 15–18 June 2010; ACM: New York, NY, USA, 2010; pp. 151–164. [CrossRef]
53. Patel, R. Upper: 14 Best Walking Route Planner Apps for Planning Your Routes. Available online: https://www.upperinc.com/blog/best-walking-route-planner-apps/ (accessed on 9 January 2024).
54. StreetWise. *StreetWise Responder for Smartphones*. 2024. Available online: https://www.streetwisecadlink.com/products/responder (accessed on 9 January 2024).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

applied sciences

MDPI

Article

Bump Feature Detection Based on Spectrum Modeling of Discrete-Sampled, Non-Homogeneous Multi-Sensor Stream Data

Haiyang Lyu, Qiqi Zhong, Donglai Jiao * and Jianchun Hua

Department of Geographic Information Science, School of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing 210023, China; hlyu@njupt.edu.cn (H.L.); 1022173217@njupt.edu.cn (Q.Z.); b21080612@njupt.edu.cn (J.H.)

* Correspondence: jiaodonglai@njupt.edu.cn

Abstract: Roads are the most heavily affected aspect of urban infrastructure given the ever-increasing number of vehicles needed to provide mobility to residents, supply them with goods, and help sustain urban growth. An important indicator of degrading road infrastructure is the so-called bump features of the road surface (BFRS), which have affected transportation safety and driving experience. To collect BFRS, we can collect discrete-sampled, non-homogeneous multi-sensor stream data. We propose a BFRS detection method based on spectrum modeling and multi-dimensional features. With the sampling rate of GPS at 1 Hz and a gyroscope and accelerometer at 100 Hz, multi-sensor stream data are recorded at three different urban areas of Nanjing, China, using the smartphone mounted on a vehicle. The recorded stream data captures a geometric feature modeling movement and the respective driving conditions. Derived features also include acceleration, orientation, and speed information. To capture bump features, we develop a deep-learning-based approach based on so-called spectrum features. BFRS detection experiments using multi-sensor stream data from smartphones are conducted, and 4, 14, and 17 BFRS are correctly detected in three different areas, with the precision as 100%, 70.00%, and 77.27%, respectively. Then, comparisons are conducted between the proposed method and three other methods, and the F-score of the proposed method is computed as 1.0000, 0.6363, and 0.7555 at three different areas, which hold the highest value among all results. Finally, it shows that the proposed method performs well in different geographic areas.

Keywords: bump feature of road surface; spectrum modeling of multi-sensor stream data; discrete-sampled and non-homogeneous dataset; movement feature

1. Introduction

With the rapid advancement of smart sensors and the influx of crowdsourced data, geospatial information for road surfaces has become increasingly accessible, crucial for the development of smart cities and, more specifically, for maintaining and optimizing road assets [1,2]. Roads, a vital component of urban infrastructure [3–5], endure significant wear due to growing traffic volumes, resulting in various surface irregularities such as bumps, cracks, potholes, and speed breakers—collectively known as bump features of the road surface (BFRS) [6,7]. Effective collection and analysis of BFRS data are essential not only for maintaining road quality but also for ensuring transportation safety and enhancing the driving experience in urban environments [8,9].

Vision-based methods, widely used for their ability to capture road surface conditions through cameras, employ techniques such as computing gradients and Haar features from video frames and images to detect BFRS [10,11]. These methods, alongside advanced machine learning techniques for pothole and crack segmentation [12], and the use of mono and binocular stereo cameras for depth imaging [13–15], form the core of intuitive vision-based BFRS detection. Recent developments in deep learning have significantly improved

these detection methods, tailoring them more effectively to the nuanced challenges of BFRS identification [16]. However, these approaches are susceptible to issues like variable lighting, data noise, and obstructions caused by environmental conditions [10,11]. Alternative approaches using accelerometers and GPS have been developed to monitor road surface anomalies directly from driving data, offering a robust complement to visual methods [9,17]. Furthermore, professional settings increasingly utilize advanced sensor technologies like LiDAR to capture precise geometric details of road surfaces, identifying BFRS through deviations from standard models [18–20]. Additionally, some methods leverage the combination of acceleration sensors, GPS, and auxiliary sensors to capture vehicle vibrations during travel, indirectly detecting BFRS by analyzing unusual sensor data patterns [3,21]. The integration of Internet of Things (IoT) technologies further enhances these methods by facilitating extensive data collection from a network of sensors, thereby creating multi-source datasets essential for addressing complex BFRS detection challenges [3,8,22]. Multi-sensor setups, particularly those incorporating smartphone technologies, offer substantial advantages over traditional methods by enhancing robustness and data diversity. The widespread adoption of smartphones has transformed these devices into powerful tools for crowdsourced data collection on BFRS, which can improve the efficiency of BFRS detection operations and combine user convenience with extensive data coverage [3,7,17,23–25]. Despite these advancements, challenges remain, including (1) managing discrete-sampled, multi-sensor stream data from smartphones in randomly positioned vehicles, (2) limited dataset coverage, and (3) the complexity of detecting and representing BFRS from non-homogeneous datasets collected under varied driving conditions.

To address these challenges and detect BFRS using the widely used motion sensor, this article concentrates on multi-sensor stream data obtained from smartphones positioned in arbitrary poses. The following was carried out: (1) modeling BFRS and augmenting the training dataset with movement features; (2) generating spectral features through fast Fourier transform (FFT) and introducing a deep learning-based BFRS detection method utilizing non-homogeneous, discrete-sampled, multi-sensor data; and (3) representing BFRS using a weighted clustering method. Rather than the numerical measurement of acceleration data from multiple smartphones or varying vehicle types, the contributions of this article are as follows: (1) the modeling of vehicle movement that can deal with the BFRS in different frequencies and driving conditions, along with the augmentation of the training dataset; (2) the representing and detecting of BFRS in account with the noisy sampling; and (3) the presentation and integration of BFRS locations based on the weighted clustering method. To evaluate our approach, we conduct experiments at different geographic areas in the urban area of Nanjing, China. A part of the dataset is used as the training dataset, and others are used as the evaluation dataset; then, comparisons are conducted between the proposed method and three different methods.

2. Related Work

As an important factor for transportation safety and driving experience, BFRS detection has long been the heated point for years [3,6,11,16]. Hence, based on the sampling device, related research can be categorized into intuitive vision-based methods and sensor-based methods.

In the intuitive vision-based method, BFRS is collected directly by the camera through images or videos [10,11,13,26]. Images or videos can be collected by specially designed cameras or smartphones during the movement on the road surface. Then, collected images or videos are resampled as a single image with a selected time interval, and then gradients, edges, Haar feature, and some other similar features are computed for BFRS. These image-based techniques are well demonstrated by several recent studies, each employing different methodologies to detect and analyze road surface conditions. Li et al. [27] focused on detecting cracks in road surface images using a Back Propagation Neural Network (BPNN). After successfully identifying the cracks, they utilized image processing techniques to

extract the related geometric shapes of each crack, demonstrating the practical application of neural networks in road maintenance. Following a somewhat similar approach but with different tools, Azhar et al. [28] applied histograms of oriented gradients (HOG) to capture textural features of the road surface. They fed these HOG features into a Naïve Bayes classifier to detect potholes. Once identified, the potholes' geometric information was extracted using the graph-cut image segmentation method, highlighting the integration of machine learning and advanced image processing techniques. Expanding on these methods, Wang et al. [10] computed the Haar feature of each sampled image and trained a classifier using the Adaboost algorithm to detect cracks on the road surface. This approach emphasizes the robustness of feature-based machine learning algorithms in identifying specific types of road damage. Ouma et al. [29] collected RGB images of road surfaces along with the GPS sensor to detect the accurate geometric information of potholes. The RGB image is first enhanced as the gray image, transformed and filtered using the 2D discrete wavelet transformation algorithm, and BFRS are clustered using the fuzzy-c means method. Finally, BFRS are morphologically reconstructed and compared with ground truth objects for validation. In addition, based on photogrammetry or computer vision theory, a mono camera can also detect the generated depth image using the binocular stereo camera. There are various innovative approaches using both image processing and machine learning techniques. Xiao et al. [14] developed a hybrid method that fuses monocular images with LiDAR point clouds to detect road surfaces in diverse scenarios. They utilized classifiers for both images and point clouds based on boosted decision trees and integrated the classification results using Conditional Random Fields (CRFs). Building on the use of imaging for detection, Zhang et al. [13] constructed a pothole detection system based on stereovision. By calculating distances from stereo camera-collected image pairs and utilizing auxiliary GPS data, they were able to precisely determine the size, volume, and position of each pothole, demonstrating the effectiveness of integrating stereo vision with geospatial data. Further exploring stereo imaging, Fan et al. [15] detected potholes using stereo images mounted on a vehicle. They employed random sample consensus (RANSAC) to mitigate the effects of potential outliers during the 3D road surface reconstruction, transformed dense disparity maps to enhance the distinction between damaged and undamaged road areas, and used threshold selection for BFRS detection. These methods underscore the growing influence of deep learning technologies in improving the accuracy of detection systems. Continuing this trend, Kulambayev et al. [16] developed a real-time road surface detection model using a Mask Region-based Convolutional Neural Network (Mask R-CNN), trained and validated with images manually labeled using Microsoft Visual Object Tagging Tool (VoTT). This approach represents a significant leap forward in applying deep learning to road surface assessment, providing high accuracy and real-time processing capabilities. Jenkins et al. [12] proposed pixel-wise road crack segmentation using U-Net, another kind of CNN model, and made a comparison with traditional machine learning algorithms, such as SVM and KNN. A more detailed comparison of vision-based methods can be found at Ma et al. [11], along with some related BFRS-detecting datasets. However, since the collected image or video can be affected by the light, noise samplings, and obstructed sampling, especially in the environment with poor ambient light, or BFRS is blocked on the ground by turbid water and other objects, it becomes difficult to detect BFRS for these vision-based methods.

In the sensor-based method, BFRS is captured using different kinds of sensors, such as LiDAR, accelerometer, gyroscope, GPS, and related auxiliary sensor [9,19–21], with the assumption that abnormal changes caused by BFRS can be recorded during the driving period. One way to collect BFRS is by utilizing the LiDAR sensor, which scans the road surface and detects BFRS based on abnormal changes between the point cloud-fitted surface and typical road surface. Building on the high-accuracy methods described by Siegemund et al. [18], which utilize a 3D point cloud from stereo cameras for detecting and reconstructing curbs and road surfaces, there is a shift in the approach with Sharma et al. [30]. It employed an ultrasonic sensor to monitor road surfaces, using the Dynamic Time Warping (DTW) technique for image processing to match and locate different kinds

of BFRS via a GPS sensor. While both methods provide precise measurements of the shape and type of BFRS, they also entail significant costs due to the professional tools required. An alternative approach involves the use of specially designed IoT sensors to collect BFRS data. Mednis et al. [31] proposed a road surface monitor method using the on-board vehicular sensor system, which was composed of a regular PC computer, low-cost microphone sensor, and GPS receiver. Potholes were detected based on selected thresholds for the microphone sensor recordings. In addition, sensor recordings can be uploaded to the web server through IoT networks [2,4], and BRFS is detected based on multi-source datasets. These methods detect BFRS based on the vibration of vehicles during the driving period, with an acceleration sensor, GPS sensor, and related auxiliary sensor. In recent years, smartphones equipped with multiple sensors have emerged as a popular tool for collecting road condition data. Li et al. [3] utilized this technology to develop a road roughness assessment method. By leveraging the built-in GPS receiver and accelerometer in smartphones, they gathered a spatial series of geo-referenced vertical accelerations of the road surface. They further enhanced this approach by creating a mobile crowdsensing system, which integrates data from multi-source smartphones and a web-based server. This system computes the International Roughness Index (IRI) for each road segment to assess surface quality. Building on this foundation, Rajamohan et al. [8] conducted experiments and developed a prototype system named MAARGHA, designed to estimate the condition and surface type of road surfaces. This work is part of a broader trend in which researchers such as Yi et al. [6], Allouch et al. [32], Mihoub et al. [33], and Mednis et al. [24] have also contributed, each exploring similar methodologies to evaluate and improve road surface assessment using advanced sensor technology and innovative data integration techniques. Zang et al. [34] mounted the smartphone on the bicycle and collected the stream data of the accelerometer and GPS sensor, then BFRS were detected based on abnormal changes using the selected threshold. However, it is the result based on the direct detection from acceleration, which can be affected by the noise of the original recordings. Different from the previous work, Li et al. [7] applied the continuous wavelet transform (CWT) to extract features of acceleration time series using the collected multi-sensor data from smartphones. It took the spatial transformation for the three-axis acceleration dataset, computed the coefficient matrix of CWT, and detected the BFRS based on the threshold of the coefficient. There are also some BFRS detection methods based on multi-sensor data of smartphones that utilize machine learning methods, such as random forest, SVM, KNN [23,25,35], or even the deep learning methods [36,37]. To encode the BFRS feature from collected multi-sensor recordings from smartphones, Chen et al. [17] designed the vibration detection system based on smartphones, accelerometers, and cameras, installed them in the vehicle, conducted the short-time Fourier transform to acceleration, and detected BFRS using the CNN. However, the acceleration recordings were collected at a specific speed with fixed multiple sensors, which limited the feasibility of the application. As described from the above methods, the sensor-based method can collect the actual road conditions and detect BFRS without the limitations of intuitive vision-based BFRS detection methods, with the advantage of multi-sensors and crowd-sourced smartphone datasets. The utilization of non-professional multi-sensor smartphones offers significant advantages, such as accessibility and crowd-sourced data collection, enabling widespread monitoring of road surface conditions. However, the inherent noise in data collection, stemming from varied sensor quality and the randomness of smartphone orientation, presents substantial challenges. These factors lead to non-homogeneous, discrete-sampled stream data that vary under different driving conditions. Feature encoding and detection are crucial in this context as they can effectively handle the uncertainty and variability of data quality, extracting meaningful patterns and insights from the noisy, complex datasets collected by smartphones.

To deal with the non-homogeneous, discrete-sampled multi-sensor stream data from smartphones, the bump feature is analyzed based on the spectrum model; BFRS is detected using different feature channels and is further represented as cluster centers. In addition,

the specific deviation and computation of IRI are not going to be conducted in this article, as are the shape and measurement accuracy of BFRS, since the multi-sensor stream data from the crowd-sourced smartphone device just has limited accuracy and cannot be taken as the replacement of professional surveying tools.

3. Methodology

Based on the analysis of related studies, there are intentions for designing the BFRS detection method using smartphones. With the multi-sensor stream data collected at three different urban areas in Nanjing, China, experiments were conducted based on the proposed method. This section introduces the methodological framework and implementation details for BFRS detection. Different from the traditional IoT sensor, it does not extensively describe the data acquisition and measurement process across different sensors or vehicles. Instead, it focuses on acceleration, orientation, and GPS data acquired from smartphones' built-in motion sensors. Here, smartphones, placed in vehicles in arbitrary positions, record the vehicle's location, orientation, and acceleration changes during travel. This allows for the analysis of trajectory records to extract vehicular movement features on the road, such as speed and direction of movement. Subsequently, spectral analysis techniques [38] are employed to represent the spectral features of the acceleration sampling data, which does not rely on direct numerical changes in acceleration to measure BFRS due to inherent sensor noise. Additionally, by simulating different vehicular movement situations based on these movement features, data augmentation is applied to the limited neural network training dataset to enhance the extraction of BFRS and improve the accuracy of the neural network. Finally, the results detected by the CNN [39] are integrated using a mean-shift-based weighted clustering method [40] to represent BFRS.

3.1. Motivation and Background

Smartphones currently contain a large array of sensors. Using smartphone sensor data, particularly acceleration data for BFRS detection, is akin to "listening" to road surface changes through an accelerometer, much like how a song records information about its environment. However, unlike voice signals, where a speaker's speed remains relatively stable, vehicle speed and acceleration fluctuate in real-time according to actual road conditions. This variability can cause the same signal to manifest in multiple ways. Consequently, one of the innovations of the proposed method is on analyzing and modeling the relationship between BFRS characteristics and vehicle speed and on enhancing features based on this relationship to accurately capture BFRS information under various speed conditions. The following are relevant for our use case: (1) the GPS sensor to collect two-dimensional location data $P(x, y)$; (2) a gyroscope sensor to collect 3-axis orientations $Ori(o_x, o_y, o_z)$; and (3) the accelerometer sensor to collect 3-axis accelerations $Acc(a_x, a_y, a_z)$. This multi-sensor data are discretely collected at different frequencies and time stamps as a data stream S denoted by (1).

$$S\{P(x, y), Ori(o_x, o_y, o_z), Acc(a_x, a_y, a_z)\} \quad (1)$$

When the vehicle passes an object with the bump of the road surface, abnormal shakes should appear, and they can be recorded by the smartphone placed or mounted in a vehicle. Data stream S captures the state of the road surface. Our goal is to detect BFRS using these data. Challenging issues are that the quality of smartphone sensors varies across devices (denoted as S_{noise}), smartphones are arbitrarily placed, and driving behaviors vary (denoted as D_{speed}), as denoted by (2). All these conditions lead to a non-homogenous multi-sensor data stream, making it difficult to extract BFRS.

$$BFRS = \text{Detect}(S_{noise})\ s.t.\ discrete\ \&\ non\text{-}homogenous \quad (2)$$

In this article, smartphones are used as sources for road surface data collection. The sensors within the smartphone are highly integrated and encapsulated within the device. However, compared to professional surveying equipment, the accuracy of sensor data from

smartphones is limited. Therefore, to obtain road surface information, the following two issues must be analyzed from both macroscopic and microscopic perspectives:

(1) Smartphone installation/placement in macroscopic aspect: The sensors in smartphones are integrated and fixed on the mainboard, meaning the spatial attitude information of the smartphone is consistent with the accelerometer and orientation sensor data. Thus, it is possible to transform the acceleration records at any moment into an initial posture perpendicular to the road surface along the z-axis using the orientation sensor. This means the smartphone can be placed in any orientation within the vehicle as per the method described in this article.

(2) Multiple sensor calibration in microscopic aspect: At the hardware level, the relative positions of the sensors are fixed during integration within the smartphone. At the system level, calibrations can be made through the smartphone's system settings to calibrate between multiple sensors (for example, in iOS, settings can be conducted via "System"->"Privacy"->"Location Services"->"System Services"->"Compass Calibration" and "Motion Calibration and Distance"). Although "data flow/drift" occurs in motion sensors like gyroscopes, where errors accumulate over time, the data preprocessing algorithms provided by smartphone operating systems, such as the "sensor fusion algorithm" in the iPhone's Core Motion framework, can transform this "dirty data" into "clean data". Consequently, the collected dataset, including acceleration and orientation, is preprocessed and filtered. Additionally, before initiating data collection, the smartphone is restarted and left idle for a period to ensure the activation of the device's sensor calibration features. During data collection, the sampling frequency of the smartphone's accelerometer and orientation sensors is consistent, and the impact of the sensors operating at short intervals, such as 100 Hz, on the results is limited compared with vehicle speeds ranging from 0 km/h to 60 km/h. Therefore, linear interpolation techniques have been employed to perform time alignment for the collected data from multiple sensors.

Additionally, the calculation of the smartphone's posture within the sensor array requires reference to electronic compasses, magnetometers, and other sensors. Therefore, during data collection, the smartphone should be kept away from strong magnetic fields to ensure the accuracy of data collection.

Therefore, the first step in our process is temporal alignment for S due to the different time stamps and sampling frequencies. The acceleration timestamp is used as a reference, and stream data from the gyroscope sensor is aligned to the acceleration recordings using linear interpolation. GPS samples are aligned in a similar fashion.

The speed information is computed based on the distance and time duration between consecutive GPS position samplings (cf. Section 3.3 for details). It is assumed that under normal driving conditions (speeds of 0–60 km/h, without sudden traffic accidents, and consistent movement), changes in speed and position over a 1-s interval are continuous. Furthermore, given that GPS signals inherently possess an error greater than 5 m, employing linear interpolation for estimating GPS positions is considered reasonable. Consequently, to correlate acceleration sampling moments with GPS location data, linear interpolation is used to estimate GPS positions at each acceleration sampling moment.

Using the smartphone's recorded 3D orientation information, spatial transformations are performed on the corresponding acceleration data at each moment. This process converts the arbitrarily oriented 3D acceleration data into acceleration information that is parallel to the xy plane and perpendicular to the z-axis of the road surface, thereby accurately representing the vertical acceleration relative to the road surface and the horizontal changes in vehicle movement at each moment. The proposed method implements the coordinate system rotation using the Direction Cosine Matrix (DCM) method. At each sampling moment of acceleration, linear interpolation is utilized to temporally align sensor attitude data. Using the three-dimensional posture data corresponding to that moment, a DCM is established. This matrix is then multiplied with the corresponding tri-axial acceleration data to achieve spatial transformation, yielding acceleration data with the z-axis perpendicular to the road surface and the xy-axis parallel to it. The 3-axis acceleration

Acc is affected by vehicle acceleration. Hence, the spatial pose of the accelerometer is reoriented by the timestamp-aligned orientation recordings $Ori_{aligned}$, and the transformed acceleration $Acc_{trans}(a_x, a_y, a_z)$ consists of a z-axis acceleration that is always vertical to the ground and x-axis and y-axis accelerations parallel to the ground (cf. (3)).

$$Acc_{trans}(a_x, a_y, a_z) = \text{SpatialTransform}\left(Acc, Ori_{aligned}\right) \quad with\ a_z \perp ground\ \&\ (a_x, a_y) \parallel ground \tag{3}$$

With the preprocessed multi-sensor stream data S, BFRS can be detected through the following steps: (i) preprocess the dataset, (ii) study the relations between driving speed and BFRS, augment the training dataset, (iii) compute the spectrum of BFRS using FFT algorithms, and (iv) extract the specific bump feature according to the modeling result; then, (v) create a deep learning model (CNN), train and validate the model, and (vi) collect BFRS in selected areas; and (vii) finally, represent the BFRS detection result based on the weighted clustering method. The BFRS detection workflow is also depicted in Figure 1.

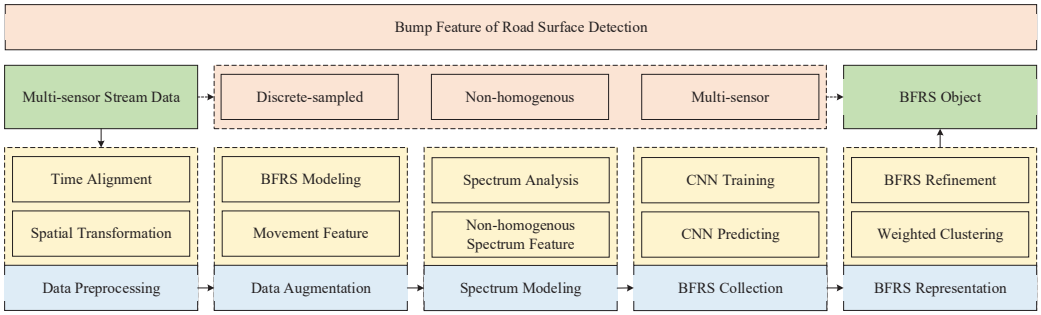


Figure 1. The BFRS detection workflow.

3.2. Model BFRS and Analyze Driving Behavior

In this work, BFRS is defined as situations/road surface conditions, including speed breakers and transverse gutters, potholes, and cracks distributed on the road surface that cause abnormal changes in the recorded data stream, as shown in Figure 2. Taking a speed breaker shown in Figure 2a as an example, vehicle A is passing the BFRS B with speed v. The length (between front wheels and rear wheels) of the vehicle and BFRS is l_A and l_B, and the accelerometer sampling frequency is h. Hence, the sampling period for object B can be denoted as (4).

$$\Delta t = \frac{l_B}{v} \tag{4}$$

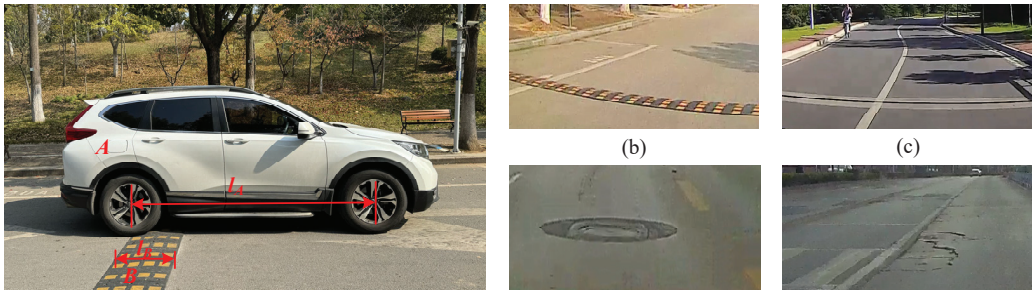


Figure 2. The BFRS detection process and different kinds of BFRS. (**a**) The BFRS detection process, (**b**) speed breaker, (**c**) transverse gutter, (**d**) pothole, and (**e**) crack.

However, based on Nyquist's theorem, a continuous signal has a maximum frequency of f_{max}, and to fully reconstruct this signal, the sampling frequency f_s must be at least twice the maximum frequency of the signal. Mathematically, it can be represented as $f_s \geq 2f_{max}$.

Therefore, the minimum frequency H to catch the bump feature can be denoted as shown in (5).

$$H = \frac{2}{\Delta t} \tag{5}$$

It can be observed that the faster the speed v or the smaller the BFRS length l_B, the higher minimum frequency H is needed. For example, in a vehicle with a length l_A = 2.7 m, a length l_B = 0.6 m, and a driving speed v ranging between 1 km/h and 60 km/h, the optimal minimum sampling frequency H should be larger than 0.93 Hz, and the frequency of BFRS ranging between 0.93 Hz and 55.56 Hz, based on (4) and (5).

Accordingly, in this work, the sampling frequency of the accelerometer and gyroscope is set at 100 Hz. On the other hand, suppose a vehicle is driving at the speed v = 30 km/h and the length l_B is 0.6 m, then the BFRS recorded frequency is 13.98 Hz. However, the length l_A between front and back wheels is 2.7 m; there can be a feature with a frequency of 3.09 Hz detected from the recordings since two bump features are recorded by front wheels and back wheels, i.e., the BFRS is recorded twice, and a new signal is generated.

3.3. Speed Computation and Data Augmentation Based on Movement Feature

BFRS also causes acceleration vertical to the road surface and a decrease in driving speed. Hence, the acceleration along all 3-axis should be taken into consideration to detect BFRS. The speed is computed using consecutive GPS samples as denoted in (6).

$$v_i = \sqrt{(x_{i+1} - x_{i-1})^2 + (y_{i+1} - y_{i-1})^2} * H_{GPS}/2 \tag{6}$$

In (6), the average speed v_i is taken as the instant speed of $P_i(x_i, y_i)$, and it is computed based on the distance between $P_{i-1}(x_{i-1}, y_{i-1})$ and $P_{i+1}(x_{i+1}, y_{i+1})$, along with the sampling frequency H_{GPS}. In addition, since the accelerometer and GPS sampling frequencies are usually different and BFRS are usually detected based on the time stamp of the accelerometer, the computed speed needs to be interpolated to align with the accelerometer time stamp. Here, the linear interpolation method is applied, as denoted in (7).

$$v_k = a(v_k - v_i) + b,\ v_k \in D_{speed} \tag{7}$$

In (7), v_i is the computed average speed at point Pi in (6); v_k is the speed interpolated for time stamp k; and a and b are coefficients of the linear interpolation function. When aligning the speed with the accelerometer time stamp, the speed v_k is computed using the precomputed speed v_i and time stamp k in sequence.

The movement feature refers to the speed and direction of the vehicle, which is recorded in the multi-sensor stream data by different driving behaviors. Based on the modeling result and movement feature, recordings containing BFRS are manually selected. However, there are only limited bump features in the training dataset. The intuitive solution is to simulate the driving behavior and augment the dataset based on the movement feature. In our work, two kinds of data augmentation methods, as shown in Figure 3, are used.

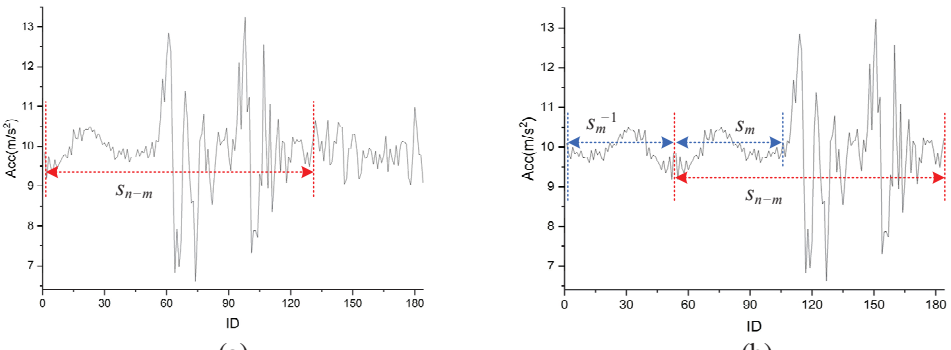


Figure 3. *Cont.*

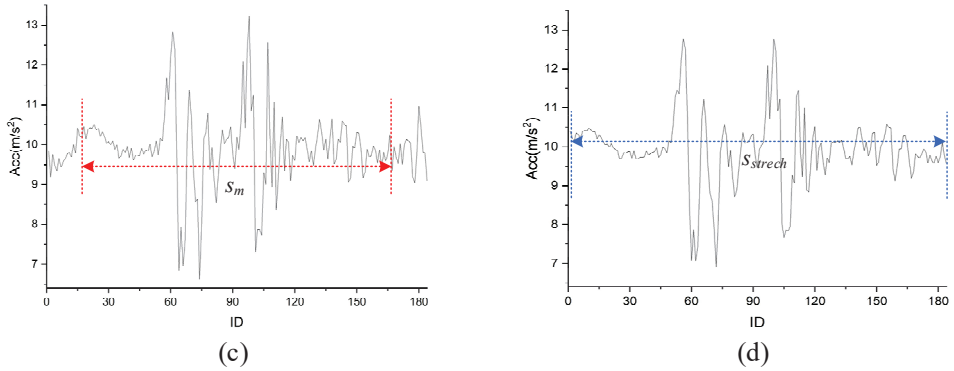


Figure 3. Data augmentation based on movement feature. (**a**) To be augmented by driving direction. (**b**) Augmented after driving direction. (**c**) To be augmented by driving speed. (**d**) Augmented after driving speed.

3.3.1. Data Augmentation Based on Driving Direction

In the context of BFRS within trajectory recordings, a recording s is composed of n samples, denoted as $s\{r_1, \quad r_2, \quad \cdots \quad r_n \}$, with driving directions indicated by time stamps. BFRS can be sampled at arbitrary positions in these segments due to different labeling methods. Typically, BFRS is positioned at the center of the recording during the manual labeling process. However, various scenarios exist for BFRS labeling, such as labeling at the beginning or end of the recording. To simulate these different scenarios, each recording undergoes a spatial transformation with a random time seed $m \in [0, w]$, and this operation is repeated $2k$ times (with k iterations for pre-resamples and k iterations for post-resamples). To maintain the length of the resampled recordings as the original, the samples on the other side of the resampled recording are repeated and mirrored. For instance, if the recording s is pre-resampled by $n - m$, it is represented as $s\{s_{pre}, s_{n-m}\}$ (as depicted in Figure 3a). However, there is a lack of samples for the previous recordings s_{pre}. Thus, the sampling s_{pre} is first represented as s_m and then mirrored as s_m^{-1} to maintain the connectivity of the recordings. The resampled recording s is then denoted as $s\{s_m^{-1}, s_{n-m}\}$ (illustrated in Figure 3b), and the entire process is represented in (8).

$$S = \text{AugmentDirection}(S, m, 2k) m \in [0, \ w], k \in N \tag{8}$$

Considering a trajectory where the vehicle moves from r_5 to r_{10}, the augmentation process involves creating continuity in the sequence. For pre-resample augmentation, if the result lacks initial recordings (e.g., $\{r_2, r_3, r_4\}$), we mirror the existing sequence (e.g., $\{r_5, r_6, r_7\}$) to fill in the gaps. In the case of post-resample augmentation, where the recording extends beyond the original sequence (e.g., $\{r_{11}, r_{12}, r_{13}\}$), the same mirroring technique is applied to maintain continuity. This approach, though simple, ensures the original sequence's continuity and effectively simulates various labeling scenarios. The augmentation process is automated and repeated to ensure a diverse dataset.

3.3.2. Data Augmentation Based on Driving Direction

In the BFRS modeling process, it is recognized that vehicles travel at varying speeds, which influences the detection of BFRS based on recordings and movement features. While multi-sensor recordings are captured at specific times and locations, they are inherently logged at a singular speed, omitting the diversity of speed scenarios encountered in real-world driving. To address this limitation, a simulation process that accounts for varying driving speeds is necessary to accurately model BFRS under different speed conditions. Let us assume the vehicle's average speed is v, and the number of multi-sensor recordings is n. To construct more realistic scenarios, it is essential to simulate speeds v' within a range $v' \in [v_{min}, v_{max}]$. The original recording s is then modified based on the comparison between v and v'. If $v > v'$, indicating slower driving conditions, the sampling period of BFRS is extended, and the recording s is stretched. Conversely, if $v < v'$, suggesting faster

conditions, the recording s is compressed. These stretching and compressing operations are executed using up-sampling and down-sampling methods, respectively. This augmentation is repeated k times with randomly selected speeds v'. For instance, when $v > v'$, a scale factor u is calculated, and only m samples $s_m\{s_{1+m/2}, s_{n-m/2}\}$ are chosen (as depicted in Figure 3c). To maintain the same length as the original recording s, up-sampling is performed through linear interpolation, resulting in $s_{stretch}$ (illustrated in Figure 3d), interpolated with n timestamps to match the length of s.

$$S = \text{AugmentSpeed}\big(S, v, v', k\big)\ v' \in [v_{min}, v_{max}], k \in N \tag{9}$$

In both (8) and (9), the maximum duration w for the random seed m for spatial transformation is typically set as half the average data length. For example, if the average dataset length is 170, w can be set to 80, and the operation repeated 40 times ($k = 20$). Similarly, the randomly selected speed v' ranges from 0 km/h to 60 km/h, with operations also repeated 40 times ($k = 40$). Moreover, while the spectrum feature remains stable during driving direction-based data augmentation, it changes after speed-based augmentation; the frequency of BFRS increases with higher speeds and decreases with lower speeds, as denoted in (4) and (5).

3.4. Spectrum Modeling for Discrete-Sampled Recordings

With the analysis result of BFRS, the next step is representing the bump feature based on the Fourier transform and related spectrum. The Fourier transform provides insights into the frequency aspects of signals in the time or spatial domain. It shifts a signal from its inherent domain, typically time or space, to a depiction in the frequency domain. This shift involves breaking down a series of data points into its individual frequency components. The FFT is a computational technique for determining the discrete Fourier transform (DFT) of a dataset or its reverse, as denoted in (10), where $D(k)$ is the output sequence in the frequency domain, $acc(n)$ is the input acceleration sequence in the time domain, N is the number of sample points, and k is the frequency index, ranging from 0 to $N - 1$.

$$D(k) = \sum_{n=0}^{N-1} Acc(n)e^{-j(2\pi/n)kn} \tag{10}$$

Analyzing the spectral components focuses on understanding the frequencies present in sampled datasets, and this mathematical equation switches a signal from its original time or spatial representation to a version defined by frequency. According to (10), the spectrum Sf is computed using sliding window w with overlap rate e that moves along the time dimension and computes FFT in a short time period. In this article, the spectrum is applied to represent the bump feature from spatially transformed Acc_{trans}, as denoted in (11).

$$Sf = \text{SpectrumAnalysis}(Acc_{trans}, l_w, o_w, H_{bfrs}) \tag{11}$$

Since the abnormal change in BFRS is recorded twice by the front wheel and back wheel of vehicles, as depicted in Figure 2, the sliding window length l_w of the BFRS to generate spectrum is set as 170, which can cover the average abnormal change under the speed condition $v \in [1, 60]$ km/h according to (4). Moreover, the frame duration is set as 50, the hop duration is set as 5, and the specific frequency H_{bfrs} of BFRS ranges from 3 to 50 Hz, based on the BFRS modeling result in Section 3.2. Assumption that the original acceleration recording is depicted in Figure 4a–c, the spectrum computed based on FFT is depicted in Figure 4d–f. It can be observed that abnormal changes in the original recordings are distributed along the time direction from the spectrum view, and the feature of the abnormal change can be represented by frequency domain. The heavier changes in the acceleration, the brighter (represented by the yellow color, in respect to the blue color) the color in the frequency domain. Moreover, compared with one-dimensional changes in the original acceleration recording in the upper row, abnormal changes in the low row are depicted just like the magnifier in the frequency domain, and the directional non-homogenous feature can be more clearly represented.

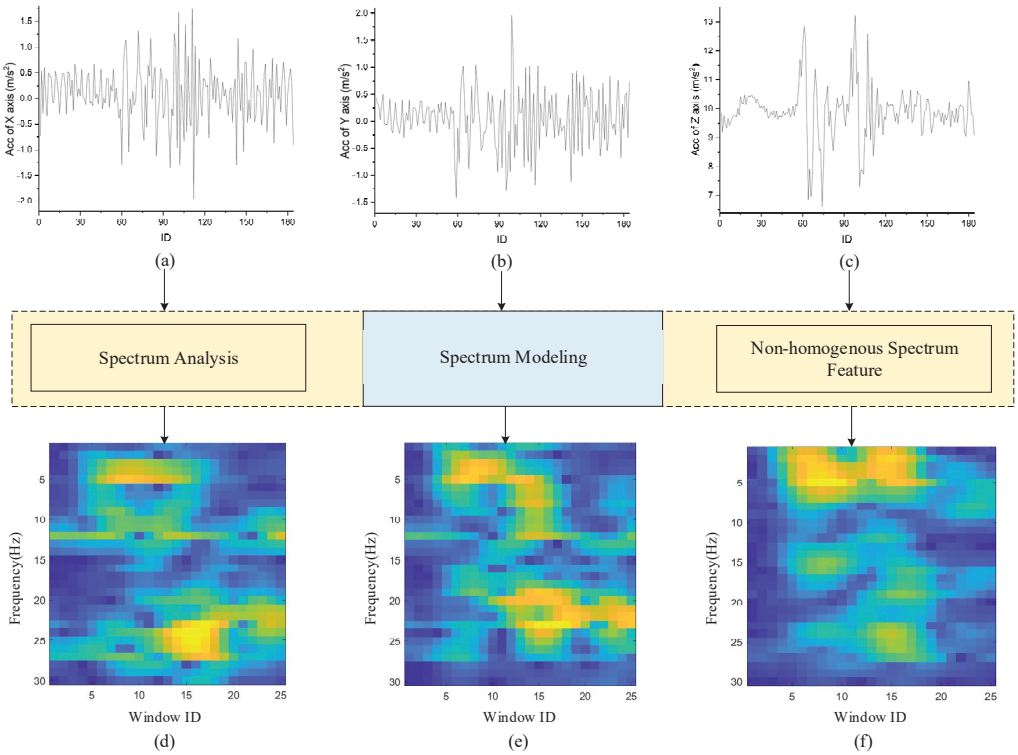


Figure 4. The non-homogenous spectrum features in different dimensions. (**a**) Acceleration on *x*-axis. (**b**) Acceleration on *y*-axis. (**c**) Acceleration on *z*-axis. (**d**) Spectrum feature on *x*-axis. (**e**) Spectrum feature on *y*-axis. (**f**) Spectrum feature on *z*-axis.

3.5. BFRS Collection Using Non-Homogeneous Spectrum Feature

As analyzed in the previous section, the abnormal changes can be observed via the *z*-axis acceleration records, while there are also similar abnormal change features in the *x*-axis and *y*-axis acceleration recordings. Therefore, the spectrum for 3-axis accelerations is computed based on (7), and four-dimensional spectra are generated, as denoted in (12).

$$\left(sf_x, sf_y, sf_z\right) = \text{MultiSpectrum}\left(Acc_{trans}\left(a_x, a_y, a_z\right)\right) \tag{12}$$

With these multi-dimensional spectra, BFRS are initially detected based on the CNN model. As a type of feed-forward neural network that learns feature engineering by itself via filter (or kernel) optimization, CNN can automatically learn meaningful features without manual engineering. By training on vast amounts of data, CNN can learn a hierarchy of features, from simple to complex, and achieve or surpass human-level performance in various tasks. Due to its exceptional performance in computer vision tasks such as image classification and object detection, CNN has become a standard technique in this domain, and it is particularly well-suited for processing matrix data. Hence, the CNN model is applied in this article to deal with the multi-dimensional spectrum data. The CNN model usually consists of the convolutional layer, activation function, pooling layer, and fully connected layer, as follows:

(1) A convolutional layer uses a set of learnable filters (or kernels) to scan input.
(2) The batch normalization layer normalizes the data on each mini-batch to accelerate the model.
(3) The activation layer is used to introduce non-linearity, with the ReLU function usually employed.
(4) The pooling layer down-samples the input data to reduce the data's dimensionality.
(5) The dropout layer turns off a subset of neurons at random during training to prevent overfitting.

(6) A fully connected layer is used for tasks like classification at the end of a CNN.
(7) The softmax layer converts a set of values into a set between 0 and 1, the sum of which equals 1.
(8) A classification layer is used for classification tasks.

In this article, a simple CNN model is designed and applied with 5 convolutional layers, 5 batch normalization layers, 5 activation layers, 4 pooling layers, 1 dropout layer, 1 fully connected layer, 1 softmax layer, and 1 classification layer. The *3*-axis acceleration data are initially encoded as a spectrum feature in dimensions of $30 \times 25 \times 3$ and converted to the size of $30 \times 25 \times 12$ by the first round of convolution operations, followed by the batch normalization. Then, the process is repeated for the other four times, with the dimensions becoming $15 \times 13 \times 24$, $8 \times 7 \times 48$, $4 \times 4 \times 48$, and $4 \times 4 \times 48$, followed by the dropout, fully connection, and softmax operations with dimension finally becoming $1 \times 1 \times 2$. Through these operations, the CNN model can "hear" the voice of the road surface and detect BFRS from the input dataset. The detailed structure is depicted in Figure 5.

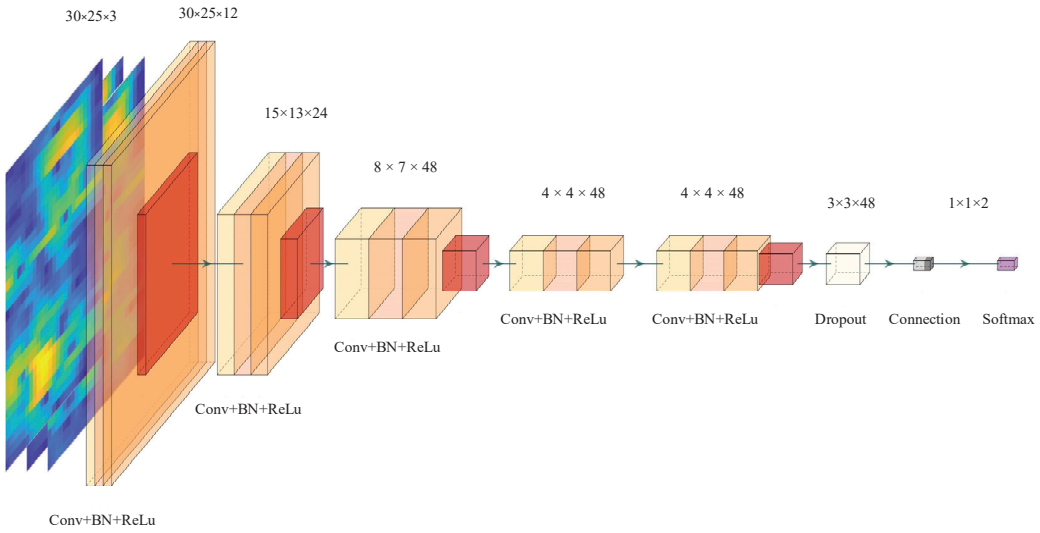


Figure 5. The BFRS detection architecture.

In Figure 5, the input three-dimensional spectra are input into the CNN model, followed by a convolutional layer with bath normalizations and an activation layer (using the ReLU function), a pooling layer (using the max pooling function), then the operation is repeated for another three rounds; finally, the BFRS are generated by the full connected layer with dropout and softmax functions. In addition, some complex CNN models can also be applied since they are just a module of the BFRS detection framework. After the training and predicting process, the $BRFS_{info}$ can be collected, as denoted in (13).

$$BFRS_{info} = \mathrm{CNN}\left(sf_x, sf_y, sf_z\right) \tag{13}$$

3.6. BFRS Representation Based on the Weighted Clustering

After the processing of the CNN model, the labels of BFRS and probability are collected as $BFRS_{info}\{BFRS_{label}, BFRS_{prob}\}$, while there is still noise and isolated detection results that have inadequate confidence and cannot be taken as proper detections. In addition, the bump feature for the same BFRS can be different, and how to deal with the non-homogeneity remains a challenge in the traditional signal processing operation. To solve the problem, the detection result is first refined based on the buffered windows that keep the consecutive confidence of the detection result. After the process, each BFRS is labeled as a segment of the original recording; hence, the second step is representing the BFRS as a single position according to the weighted clustering. The refinement and clustering process can be depicted in Figure 6.

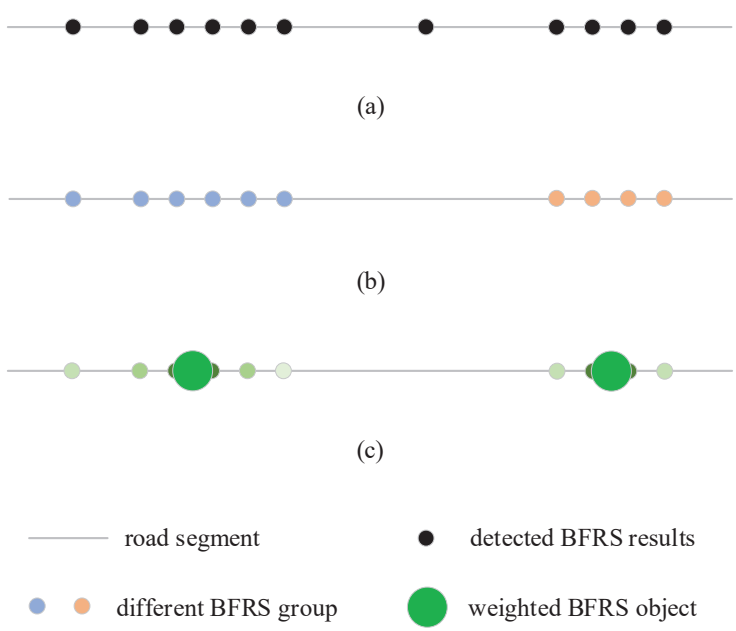


Figure 6. The BFRS representation from segment to position. (**a**) BFRS detected by CNN. (**b**) Refined and segmented as each cluster group. (**c**) Represented by weighted cluster center.

3.6.1. Refinement Using the Buffered Window

Suppose the detected result is subtracted with a buffered window w_{buf} that slides from the start time to the end of the recording, which can be denoted as $w_{buf}\{BFRS^{i}_{info}, BFRS^{i+1}_{info}, \cdots, BFRS^{i+n}_{info}\}$. Then, the refinement process is conducted for each result in w_{buf}. The detected result in w_{buf} is judged as BFRS under the following three conditions:

(1) More than half of the detected result $BFRS_{label}$ is not labeled as "road".
(2) More than 60% percent of detected results in the buffered window w_{buf} is BFRS.
(3) The probability $BFRS_{prob}$ of a detected result is larger than the probability threshold $Prob_{min}$.

The point in the buffered window is relabeled as "true" and taken as the right detected BFRS, or the detected is taken as "false", and the operation can be depicted in (14).

$$BFRS_{refined} = \text{Refine}\left(BFRS_{info}, w_{buf}, Prob_{min}\right) \tag{14}$$

In (14), the length n of the buffered window w_{buf} is usually set as 10, or some other larger value that with the high confidence of the detection result, and the probability threshold $Prob_{min}$ is set as 0.7, which is a little higher than the probability detected by the CNN model, as depicted in Figure 6a.

3.6.2. Representation Using Weighted Clustering

With the refinement process, the detected result $BFRS_{refined}$ is converted to each segment $BFRS_{group}$, which is composed of consecutive points $BFRS_{group}\{b_1, b_2, \cdots, b_n\}$. However, the BFRS is the specific position, such as the speed breaker or pothole, so the next step is to distinguish the segments $BFRS_{group}$ from each other and convert the consecutive segment into a specific object location $BFRS_{obj}$, as depicted in Figure 6b. The operations are as follows:

(1) Compute the difference index of each detected sampling point based on the index orders from the original recordings.

(2) Segment detected sampling points into each group $BFRS_{group}$ if the index difference is within the threshold value d, as denoted in (15).

$$BFRS_{group} = \text{Cluster}\Big(BFRS_{refined}, d\Big) \tag{15}$$

(3) Each detected sampling point in $BFRS_{group}$ is composed of the specific coordinate $P_i(x, y)$ and the related probability $BFRS^i_{prob}$; hence, points in each $BFRS_{group}$ can be further integrated into a single point $BFRS_{obj}$ (depicted in Figure 6c) based on the weight of $BFRS^i_{prob}$, as denoted in (16).

$$BFRS_{obj} = \text{Weight}\Big(BFRS_{group}, BFRS_{prob}\Big) \tag{16}$$

In (15), there is a key parameter d that segments the consecutive detected sampling point into each group, and it can be selected depending on the data quality of the original recordings. Usually, d can be set as 3 if too many noisy samplings are not included. While in (16), the coordinate of $BFRS_{obj}$ in each $BFRS_{group}$ is automatically computed based on each detected sampling point and the related $BFRS_{prob}$ (depicted as the graded green color in Figure 6c), and no other parameters are needed.

3.7. BFRS Detection Algorithm from Multi-Sensor Stream Data

The BFRS detection framework consists of the following four steps: Firstly, preprocess the collected discreet multi-sensor stream data from smartphones, including (1) timestamp alignment for multi-sensor recordings based on linear interpolation and (2) spatial transformation for the 3-axis acceleration data. Secondly, augment the training dataset based on the movement feature, model BFRS using the driving speed and 3-axis acceleration recordings, and represent the bump features based on spectrum modeling. Thirdly, design the CNN-based deep learning network to deal with the non-homogenous BFRS spectrum and collect BFRS. Finally, represent the BFRS detection result based on the weighted cluster method. The algorithm is depicted in Algorithm 1.

Algorithm 1 The BFRS detection algorithm

Input : $\boldsymbol{S_{noise}}$ denoted as (1) and (2)
Output: *BFRS* denoted as (2)
BFRSDection($\boldsymbol{S_{noise}}$) $\Big\{\boldsymbol{P(x,y)}, \boldsymbol{Ori}(o_x, o_y, o_z), \boldsymbol{Acc}(a_x, a_y, a_z), \boldsymbol{D_{speed}}\Big\}$
//step 1: preprocessing and spatial transformation
$\boldsymbol{Ori_{aligned}}$ = AlignOrientation($\boldsymbol{Ori}(o_x, o_y, o_z), t_{acc}$); //Align the Orientation with Acceleration
$\boldsymbol{Acc_{trans}}(a_x, a_y, a_z)$ = SpatialTransform($\boldsymbol{Acc}(a_x, a_y, a_z), \boldsymbol{Ori_{aligned}}$); //based on (3)
//step 2: data augmentation and BFRS spectrum modeling
$\boldsymbol{H_{bfrs}}$ = BFRSModeling($\boldsymbol{S_{noise}}, \boldsymbol{D_{speed}}$) //based on (4) and (5)
$\boldsymbol{S}\{\boldsymbol{P}, \boldsymbol{Acc}\}$ = AugmentDataset($\boldsymbol{Acc_{trans}}, \boldsymbol{D_{speed}}$)//based on (6)–(9)
$Sf(\boldsymbol{sf_x}, \boldsymbol{sf_y}, \boldsymbol{sf_z})$ = ComptueSpectrum($\boldsymbol{S}, \boldsymbol{H_{bfrs}}$); //based on (10)–(12)
//step 3: BFRS collection based on CNN
$\boldsymbol{para_{CNN}}$ = TrainCNN(*Sf*);
FOREACH Sf_i IN *Sf*
  W_i = SildingWindow($\boldsymbol{Sf_i}$); //slice the spectrum
  $\boldsymbol{BFRS^i_{info}}\{\boldsymbol{BFRS_{label}}, \boldsymbol{BFRS_{prob}}\}$ = PredictCNN($W_i, \boldsymbol{para_{CNN}}$); //based on (13)
END
$\boldsymbol{BFRS_{refined}}$ = **BFRSRefine**($\boldsymbol{BFRS_{info}}, \boldsymbol{w_{buf}}, \boldsymbol{Prob_{min}}$) //based on (14)
//step 4: BFRS representation based on weighted clustering
$\boldsymbol{BFRS_{group}}\{\boldsymbol{P_{BFRS}}, \boldsymbol{BFRS_{prob}}\}$ = Clustering($\boldsymbol{BFRS_{refined}}, \boldsymbol{d}$); // based on (15)
$\boldsymbol{BFRS_{obj}}$ = WeightedIntegration($\boldsymbol{BFRS_{group}}$); // based on (16)

4. Experiments and Discussion

This section focuses on the performance of the BFRS detection method. The multi-sensor stream data are acquired using smartphones with GPS, gyroscope, and accelerometer sensors. GPS locations, three-axis orientations, and three-axis acceleration datasets are collected. Smartphones are used as the source for sensor data collection, utilizing the inherent data processing techniques of the smartphone operating system. This setup effectively addresses the issue of "data flow/drift" in motion sensors (such as the gyroscope). The raw acceleration data are preprocessed using these techniques before being recorded by the MATLAB Mobile 2023 app and subsequently sent to the proposed method. Each dataset collected typically lasts no longer than 10 min. Due to the robust preprocessing capabilities of the smartphone, extra calibration for the motion sensors is not conducted during the data collection process. BFRS is modeled after timestamp alignment between the accelerometer and GPS and gyroscope sensors. Then, the direction of the three-axis accelerometer is reoriented by the three-axis acceleration information, and the spectrum model is applied to the accelerometer data. The BFRS is computed based on CNN and represented using the weighted cluster method. After that operation, the cluster result is compared with the actual BFRS positions, treating detections within 10 m of the actual BFRS as correct and those outside this range as wrong. These criteria serve as our evaluation metrics, and precision, recall, and F1-score are calculated. Based on the criteria, comparisons are conducted with different BFRS detection methods; in addition, BFRS in different areas is computed, and results are evaluated with different methods.

4.1. Multi-Sensor Stream Data Collected Using Smartphone

As shown in Figure 7, the total length computed based on GPS recordings is 6411 m with speed ranges from 0 km/h to 39.65 km/h, in an area of 0.92 km^2. The time duration of the stream data is 860 s, with 860 sampling of GPS points and 86,662 records of three-axis acceleration and three-axis orientation records. In the research area, BFRS refers to surface features that have variations in elevation greater than 3 cm and extend horizontally along the road for more than 4 cm. Due to the geographic coordinates inherent in the detection results, which differ from traditional methods of evaluating neural network models, data from a specific spatial area was selected as the training set, and data from a different area (not included in the training set) was used for testing. In Area A, the dataset is divided into a training dataset (outside the black dotted rectangle area) and a test dataset (inside the black dotted rectangle area, i.e., Area A), and the training dataset is used to train and validate the CNN model, while the test dataset is used to make validation and comparison with other related methods. The detailed statistics of the collected multi-sensor stream data are depicted in Table 1.

Table 1. Statistic of collected multi-senor stream data.

	Length	Duration	GPS Records	Acceleration Records	Orientation Records
Area A	6411 m	860 s	860	86,662	86,662

location recordings

time	latitude	longitude
09:00:06.029	32.11913	118.9237
09:00:12.638	32.11909	118.9236
09:00:19.482	32.11915	118.9235
09:00:26.066	32.11896	118.9231
09:00:32.651	32.11871	118.9231
09:00:39.309	32.11772	118.9228
09:00:45.933	32.1173	118.9227
09:00:52.557	32.1166	118.9226

acceleration recordings

time	AccX	AccY	AccZ
09:00:00.003	2.790048	−1.49404	9.513018
09:00:00.013	2.922672	−0.32932	9.309591
09:00:00.023	2.798431	−1.42878	9.522747
09:00:00.033	1.551823	−0.67525	10.46923
09:00:00.043	1.238075	−0.32857	8.856183
09:00:00.054	2.124083	−0.65953	9.256601
09:00:00.064	2.46193	0.144899	8.539592
09:00:00.074	1.624871	−0.23127	8.202942

orientation recordings

time	OriX	OriY	OriZ
09:00:00.004	−82.9393	2.802022	−12.6739
09:00:00.014	−82.9267	2.806715	−12.6555
09:00:00.024	−82.9199	2.81774	−12.679
09:00:00.034	−82.8996	2.830981	−12.6914
09:00:00.043	−82.8872	2.847044	−12.7077
09:00:00.054	−82.8742	2.867247	−12.7135
09:00:00.065	−82.8694	2.887331	−12.7041
09:00:00.075	−82.8587	2.912691	−12.7232

(a) (b)

road segment recorded trajectory BFRS training dataset

test area recorded trajectory in test area

Figure 7. The research area and related multi-sensor stream data. (**a**) Research area. (**b**) Related multi-sensor recordings.

4.2. Non-Homogeneous Analysis and BFRS Modeling

The smartphone is arbitrarily mounted or placed in the vehicle, but there is an extensive abnormal change in the direction vertical to the road surface, and the spatial transformation is necessary to convert the z-axis acceleration into the vertical direction of the road surface. With timestamp-aligned three-axis acceleration and three-axis orientation recordings, the real-time spatial transformation is applied to the three-axis acceleration. The spatial transformation for the acceleration stream data is depicted in Figure 8.

As shown in Figure 8, there is an obvious change in the three-axis acceleration before and after the spatial transformation. In Figure 8a, both x-axis, y-axis, and z-axis accelerations change during the sampling period, and the average acceleration for the three axes is 2.2278 m/s^2, -0.6017 m/s^2, and 9.4296 m/s^2, respectively. In Figure 8b, the average x-axis and y-axis acceleration becomes 0.0513 m/s^2 and -0.0533 m/s^2, which is consistent with the actual case (as the vehicle moves forward at a nearly constant speed), and the z-axis acceleration is to 9.7075 m/s^2; however, abnormal changes are concentrated on the z-axis, while there are also related slight changes in the x-axis and y-axis. It shows strong relations with BFRS after spatial transformation. Additionally, it can be observed that BFRS impacted both the vertical (z-axis) and horizontal (xy-axis) directions of the vehicle. In

the z-axis direction, the changes in acceleration are quite noticeable, typically fluctuating above ± 3 m/s^2. In the x-axis direction (assuming it is the direction of vehicle travel), a noticeable deceleration effect is also presented, although the changes in acceleration were relatively minor. However, employing a threshold method to determine BFRS can lead to significant errors due to the inherent noise in acceleration recordings. Therefore, the proposed method utilizes spectral features to detect BFRS, which also aids in mitigating the impact of acceleration noise to some extent.

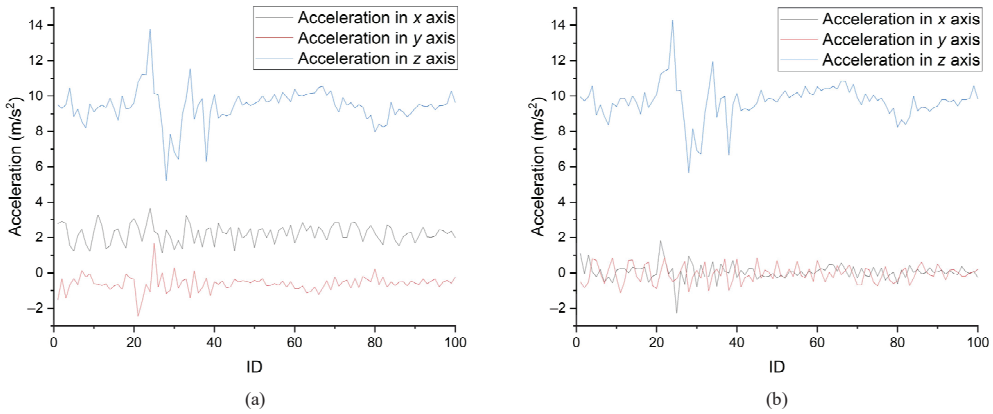


Figure 8. Spatial transformation for the acceleration data. (**a**) Before spatial transformation. (**b**) After spatial transformation.

BFRS is modeled with the relations between the geometric shape and driving speed in this research, and the BFRS training dataset is augmented based on the modeling result since the size and quality of the training dataset have a direct effect on the neural network. To have an intuitive visualization of the augmented results, the experiment is conducted on the actual collected recordings using the proposed augmentation method, as depicted in Figure 9. In the experiment, the spectrum of BFRS is computed using the FFT algorithm based on (10)–(12), with the parameter set as Section 3.3.

As seen from the experiments in Figure 9, the spectra of BFRS are computed and visualized for the acceleration of the x-axis, y-axis, and z-axis. In Figure 9a–c, it is the comparison result of data augmentation by driving direction, and the original recording in Figure 9b is pre-resampled with 33 samplings, while post-sampled with 41 samplings. In Figure 9d–f, it is the comparison result of data augmentation by driving speed; the original recording in Figure 9e is speed increased with 123% of the original speed in Figure 9d, while decreased with 85% of the original speed in Figure 9e. In Figure 9a–c, the spectrum feature keeps steady during the pre-resample and post-resample processes. However, in Figure 9d,e the same spectrum features vary under different speed conditions; the spectrum feature representation of the same BFRS is stretched or squeezed by different vehicle speeds and becomes non-homogenous. In addition, to validate the effectiveness of the data augmentation method, two different spectrum feature maps are generated based on the actual acceleration recording under speed conditions of 18 km/h and 27 km/h, as depicted in Figure 10.

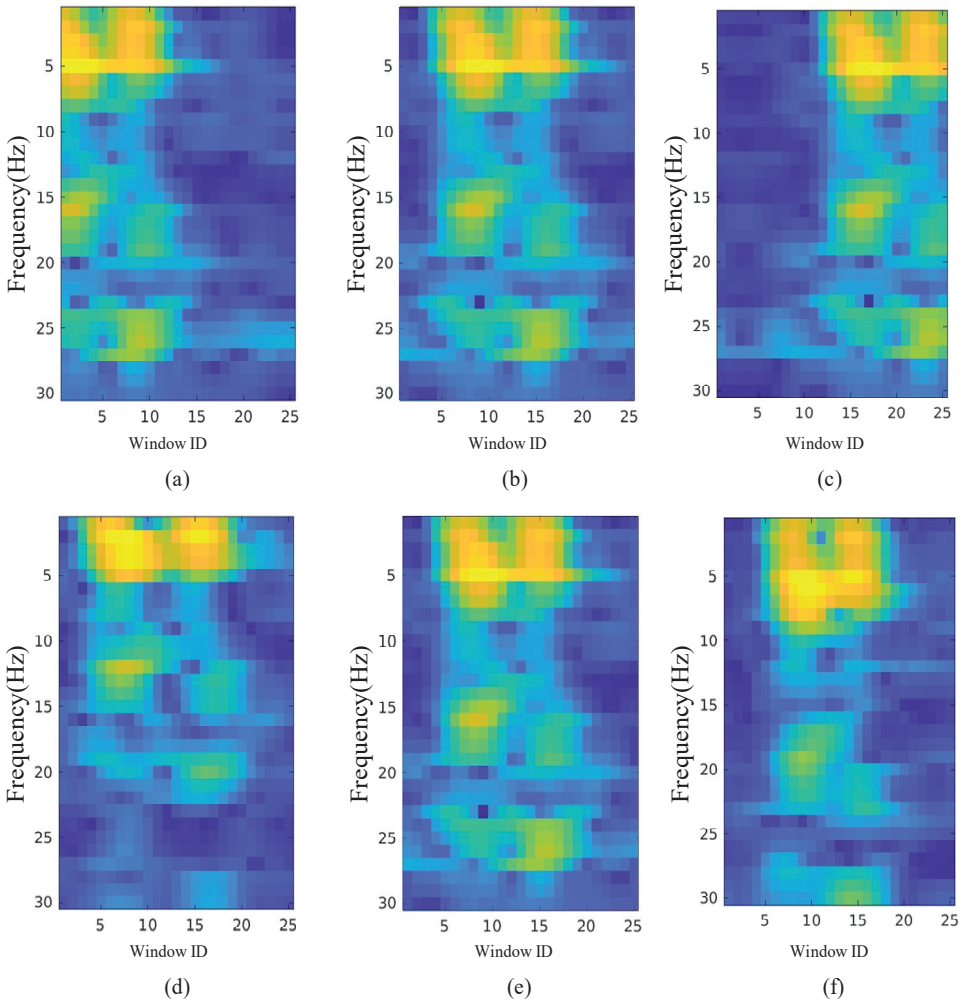


Figure 9. Comparisons of spectrum modeling results. (**a**) Augmented by driving direction: pre-resample. (**b**) Augmented by driving direction: original dataset. (**c**) Augmented by driving direction: post-resample. (**d**) Augmented by driving speed: low speed. (**e**) Augmented by driving speed: original dataset. (**f**) Augmented by driving speed: high speed.

In Figure 10, the actual acceleration recording of the same kind of BFRS is collected under different speed conditions. There is a strong saliency (the light-yellow color in Figure 10b) in the lower frequency domain when recording under the speed of 18 km/h in Figure 10a; however, the strong saliency (the light-yellow color in Figure 10d) moves to the higher frequency domain when recording under the speed of 27 km/h in Figure 10c, which is consistent with the assumption analyzed in Section 3.3. In addition, the augmentation result based on the speed direction is not further experimented with, as the spectrum feature keeps steady during the process. Despite the presence of noise in the original recordings, the method transforms one-dimensional acceleration changes into two-dimensional spectral features. By distinguishing BFRS characteristics from noise across different frequency spectrum coefficients, the proposed method effectively reduces the impact of noise on experimental results. Based on the proposed data augmentation, there are non-homogenous feature representations under different speed conditions, and the feature can be extracted and strengthened based on the spectrum modeling method, which would contribute to the classification of BFRS in the next section.

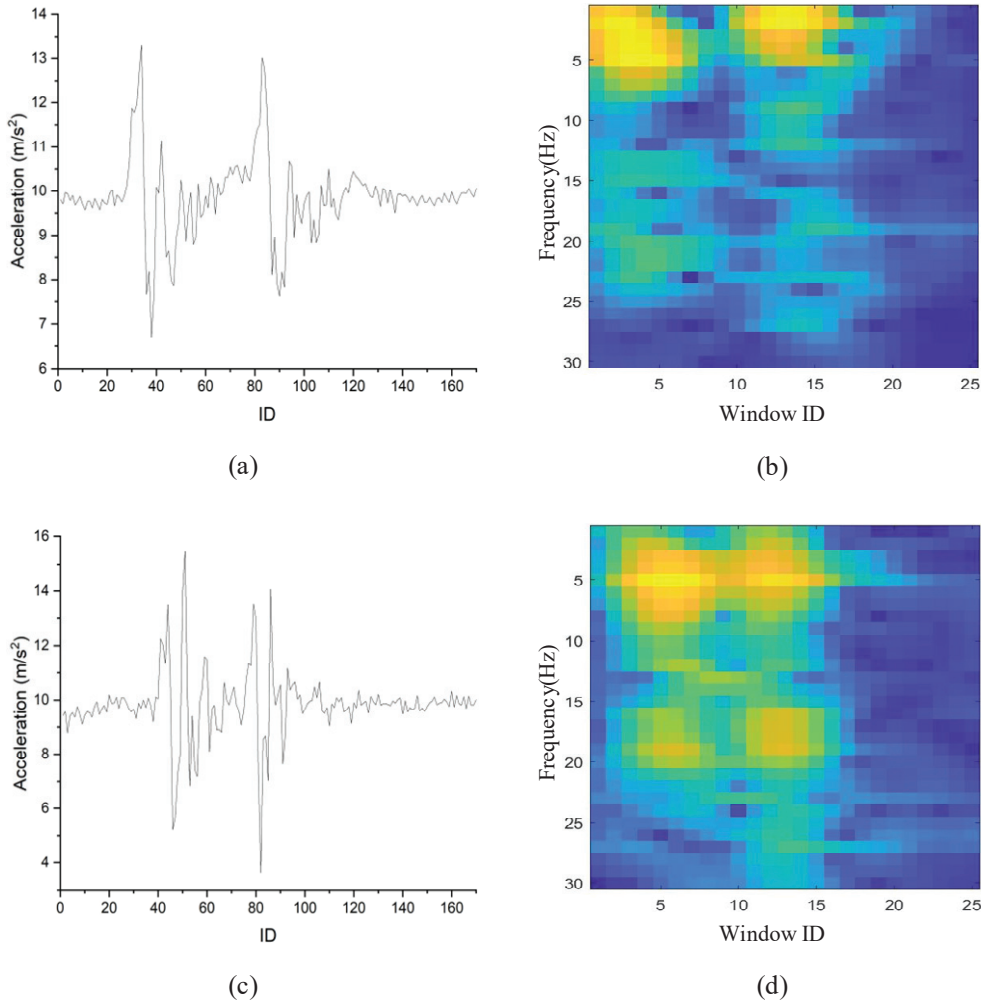


Figure 10. The actual acceleration recording and related spectrum feature. (**a**) Acceleration recording with actual speed 18 km/h. (**b**) Related spectrum feature of 18 km/h. (**c**) Acceleration recording with speed 27 km/h. (**d**) Related spectrum feature of 27 km/h.

4.3. BFRS Detection and Representation

Since BFRS is affected by the abnormal changes in the three-axis acceleration, all the spectra in three dimensions are computed and taken into the CNN model. In addition, the non-homogenous is caused by different driving speeds, as discussed in the previous section. The CNN model to detect BFRS is designed according to Section 3.3, with the spectrum in the three-axis acceleration. The training parameters are set as follows: {number of filters: 12; mini batch size: 50; learning rate: 0.0001; max epochs: 30; dropout probability: 0.2; padding mode: same}, with other parameters the same as Section 3.5. During the training process, the dataset is divided into the following two categories: 80% of datasets are taken as training datasets, and 20% datasets are taken into validate datasets. Furthermore, to distinguish the BFRS from the normal road features, 166 segments of normal road segments are randomly selected in the research area and further labeled as "background". During the training process, the detection result is validated after each epoch, and the CNN model converges. Then, the BFRS detection operation is conducted on the test dataset of Area A, with the trained learnable parameters. The multi-dimensional spectrum features of acceleration recordings are sent to the CNN model, output with the detected labels, segmented and refined as consecutive segments, and weight-clustered and represented as BFRS, as depicted in Figure 11.

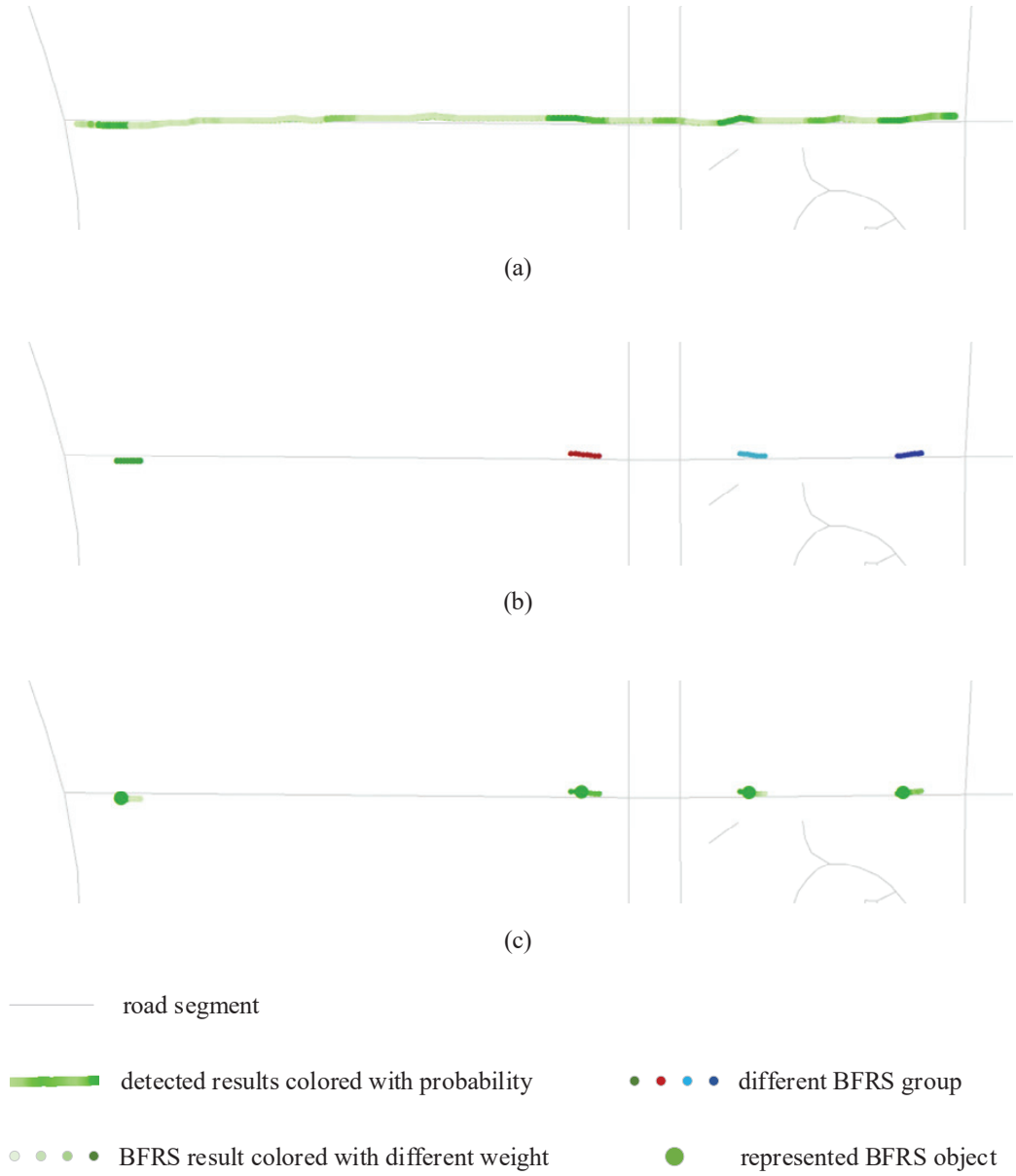


Figure 11. BFRS detection results. (**a**) BRFS detected results colored with probability. (**b**) Segmented and refined as different BFRS groups. (**c**) Represented by weighted cluster center.

In Figure 11a, the BFRS is initially detected based on the CNN model; each recording is assigned the probability of BFRS and further visualized by the graded color in greed (the darker the point, the higher the probability of the BFRS). In Figure 11b, the detection result is further segmented and refined as each segment based on (14) and (15), which are labeled with different colors. Finally, the BRFS is represented as the single BFRS in each group using the weighted clustering process based on (16). In Figure 11, it can be seen that the BFRS are dynamically generated based on the proposed BFRS detection method in Section 3.7. The proposed method primarily aims at the positioning of BFRS, but it can also represent the size of the obstacle. For the horizontal length of BFRS, it can be directly represented by the results of continuous detection, that is, without clustering the BFRS initially (as shown by the different BFRS group in Figure 11). The lateral width can be inferred from multi-source data coverage of the road surface, which helps estimate the lateral width of BFRS. The changes in z-axis acceleration from the recorded acceleration corresponding to detection results can indicate the vertical dimension (as shown by the actual acceleration recording in Figure 10).

In addition, from the acceleration data recorded, ideally, when a vehicle passes over a hump-type obstacle, it produces an upward acceleration followed by a downward acceleration. Conversely, passing over a hole results in the opposite pattern of acceleration. However, these features are not distinctly clear and are difficult to differentiate due to inherent noise in the acceleration data. Theoretically, since it uses spectrum analysis coefficients to represent BFRS features, and these coefficients are absolute values from the Fourier transform process, the directionality of accelerations caused by hump-type and hole-type obstacles is neutralized by the absolute value operation. Therefore, it is not possible to accurately capture the specific acceleration features of raised and lowered surface elements. What is more, regarding road surface irregularities, such as selected thickness and type of paint, it is possible to sample these characteristics, represent them using the spectrum features, and input them into a neural network model as part of the training set. The network is subsequently trained to learn the subtle variations and distinctions among these features, thereby enabling the detection of these types of BFRS.

4.4. BFRS Detection Comparison with Different Methods

To have an intuitive understanding of the performance of the proposed BFRS detection method, different experiments are conducted, and related comparisons are made with other methods. Firstly, it is the performance comparison between the original CNN model and the proposed BFRS detection method on the test dataset (Area A shown in Figure 7). The BFRS is detected using the original CNN model without the dataset augmented based on the movement feature analyzed in Sections 3.2 and 3.3. The experiment is reconducted based on the proposed method with the same parameter as that of the original CNN model, and the result is depicted in Figure 12.

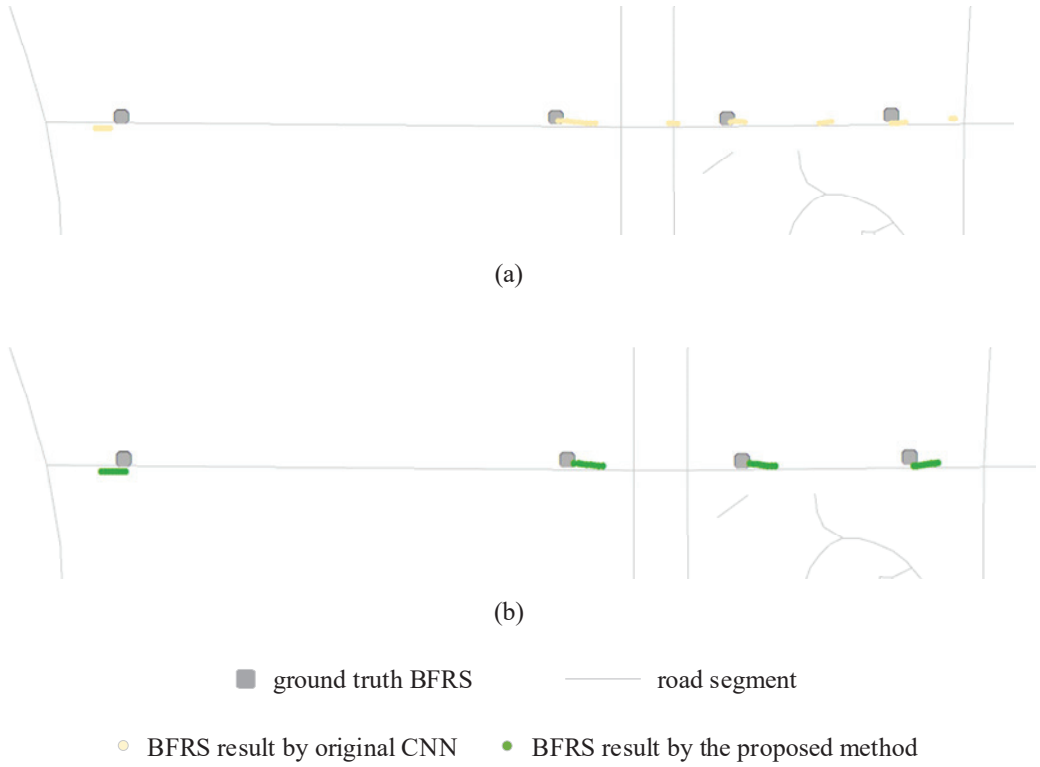


Figure 12. Comparison with and without augmentation information. (**a**) BRFS detected results without augmentation. (**b**) BRFS detected results with augmentation.

In the test dataset of Area A, there are four ground truth BFRS distributed along the road, as depicted by the gray square in Figure 12. In Figure 12a, it is the BFRS detection result based on the original CNN model [17] without augmentation information. Although four BFRS are covered by the detected BFRS groups, the results are over-detected, i.e., there

are seven BFRS groups detected and three are falsely detected. In comparison with the result in Figure 12a, the BFRS detection result in Figure 12b seems better; four BFRS are successfully covered by the detected BFRS group, and no more false-BFRS results are further detected. The over-detected BFRS result in Figure 12a is mainly caused by the insufficient training dataset, and it cannot detect BFRS in different movement situations. Hence, the proposed BFRS detection method turns out to be effective with augmentation information.

With the result in Figure 12, the BFRS detection results are further represented by the method in Section 3.6. Secondly, BFRS detection experiments are further conducted and compared with the adaptive threshold-based method [34] and the CWT-based method [7], as depicted in Figure 13.

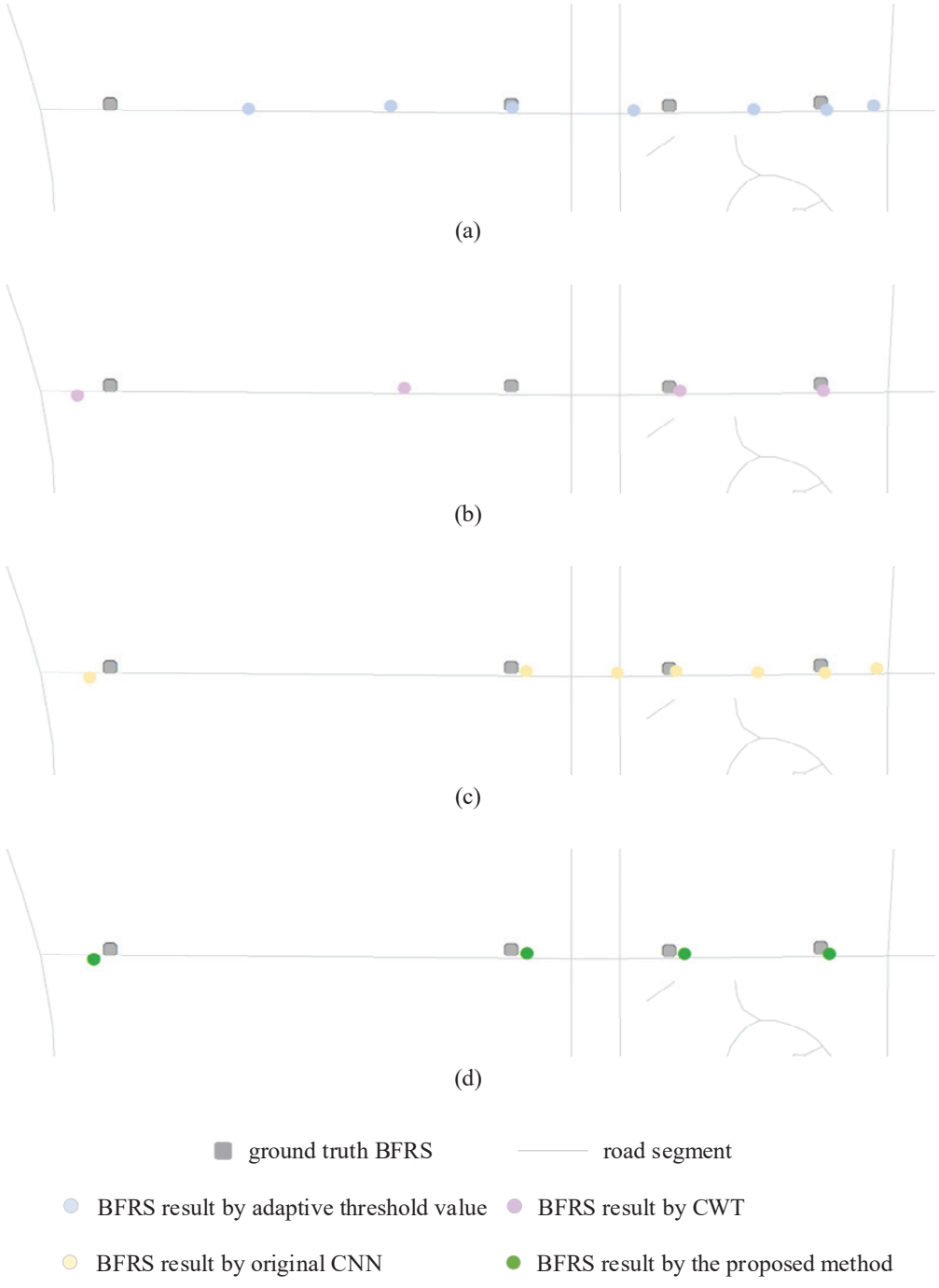


Figure 13. Comparisons between different BFRS detection methods. (**a**) Adaptive threshold value. (**b**) CWT. (**c**) Original CNN. (**d**) The proposed method.

In the result, the BFRS within the distance of 10 m is considered to be the correct BFRS, and the one out of the distance is taken as the wrong BFRS. In Figure 13a, it is the result by adaptive threshold value, computed based on the abnormal acceleration changes in its neighborhood. The first derivative of acceleration recordings is computed and taken as the abnormal change, along with the total standard deviation of the dataset. Then, the abnormal change larger than the standard deviation is adaptively selected as BFRS. Finally, seven BFRS are detected, with two correct BFRS and five wrong BFRS. In Figure 13b, the BFRS is computed based on the CWT method, with the parameters {wave name: 'db3', total scale: 256}, and the threshold value adaptively computed based on the absolute mean value of the coefficient matrix. The acceleration recording with a related coefficient larger than ten times the threshold value is selected as BFRS, and four BFRS are detected, with two correct BFRS and two wrong BFRS. Consistent with the experiment in Figure 12, the BFRS detection results of the original CNN and the proposed method are depicted in Figure 13c,d. In Figure 13c, seven BFRS are detected by the original CNN, with four correct BFRS and three wrong BFRS, while in Figure 13d, four BFRS are all correctly detected by the proposed method.

$$Precision = \frac{TP}{TP + FP} \tag{17}$$

$$Recall = \frac{TP}{TP + FN} \tag{18}$$

To evaluate the performances of different methods, precision and *recall* are computed based on (17) and (18). In the equation above, *TP* is the true positive BFRS (the true BFRS that successfully detected), *FP* is the false positive BFRS (the BFRS that falsely detected), and *FN* is the false negative BFRS (the true BFRS that are not detected). Finally, the *F-score* is further computed based on these two quality indexes, as denoted in (19), and the statistical result is depicted in Table 2.

$$F\text{-}score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{19}$$

Table 2. Statistic of the comparison results.

	Detected BFRS	**Correct BFRS**	**Wrong BFRS**	***Precision***	***Recall***	***F-Score***
Adaptive threshold value	7	2	5	28.57%	50.00%	0.3636
CWT	4	2	2	50.00%	50.00%	0.5000
Original CNN	7	4	3	57.14%	100.00%	0.7272
The proposed method	4	4	0	100.00%	100.00%	1.0000

In Table 2, the *precision*, *recall*, and *F-score* are also computed, with the number of correct BFRS and wrong BFRS. For example, the number of detected BFRS, correct BFRS, and wrong BFRS by the adaptive threshold value are 7, 2, and 5, respectively, and the related *TP*, *FP*, and *FN* can be computed as 2, 5, and 2. Hence, the *precision* and *recall* are 28.51% and 50.00% based on (17) and (18), and the *F-score* is computed as 0.3636. The *precision*, *recall*, and *F-score* for the CWT method {*TP*: 2, *FP*: 2, *FN*: 2} are 50.00%, 50.00%, and 0.5000, respectively, and these indexes for the original CNN method {*TP*: 4, *FP*: 3, *FN*: 0} are 57.14%, 100.00%, and 0.7272. However, the *precision*, *recall*, *and F-score* for the proposed method {*TP*: 4, *FP*: 0, *FN*: 0} are 100.00%, 100.00%, and 1.0000, since all BFRS are correctly detected with no wrong result.

In the experiment result in Figure 13 and Table 2, more BFRS are detected than other methods by the adaptive threshold method, while most of them are wrong results since the threshold value cannot be adapted to every situation. It seems that the CWT method can perform well with four detected BFRS, while it also faces the similar situation of the adaptive threshold value method, and the threshold only fits limited conditions. In the original CNN method, the detection standard can be learned from the training dataset,

with four corrected BFRS detected. However, the performance still depends on the training dataset, and it performs not so well since there are three wrongly detected BFRS. Finally, four BFRS are successfully detected with the highest *F-score* by the proposed method with a movement–feature-augmented training dataset. As shown in the experiments, the proposed method achieved superior BFRS detection results, with the detected positions closely aligning with the ground truth BFRS locations. Unlike the CWT method and adaptive threshold method, the proposed method does not require the setting of specific parameters. Furthermore, compared to the original CNN method, the proposed method enhances the data by considering the characteristics of vehicle movements and multi-sensor recordings, resulting in superior experimental outcomes. Hence, it can be concluded that the BFRS detection result computed by the proposed method outperforms other results in the experiment. In addition, variations in tire or suspension systems across different vehicles may impact the BFRS detection results, and BFRS might not be adequately captured just by a single vehicle type. However, it assumes general conditions, and the credibility of BFRS detection results can be further improved through the integration of multi-source data.

4.5. BFRS Detection and Comparison in the Different Areas

To assess the capability of the proposed BFRS detection method, multi-sensor stream data in different areas are collected using smartphones, and experiments are conducted based on the trained deep learning model in the previous section. The sampling rate for the GPS sensor is 1 Hz, and that for the gyroscope and accelerometer is 100 Hz. The research area and discrete-sampled GPS stream data are depicted in Figure 14a,b, with augmentation information visualized by graded color.

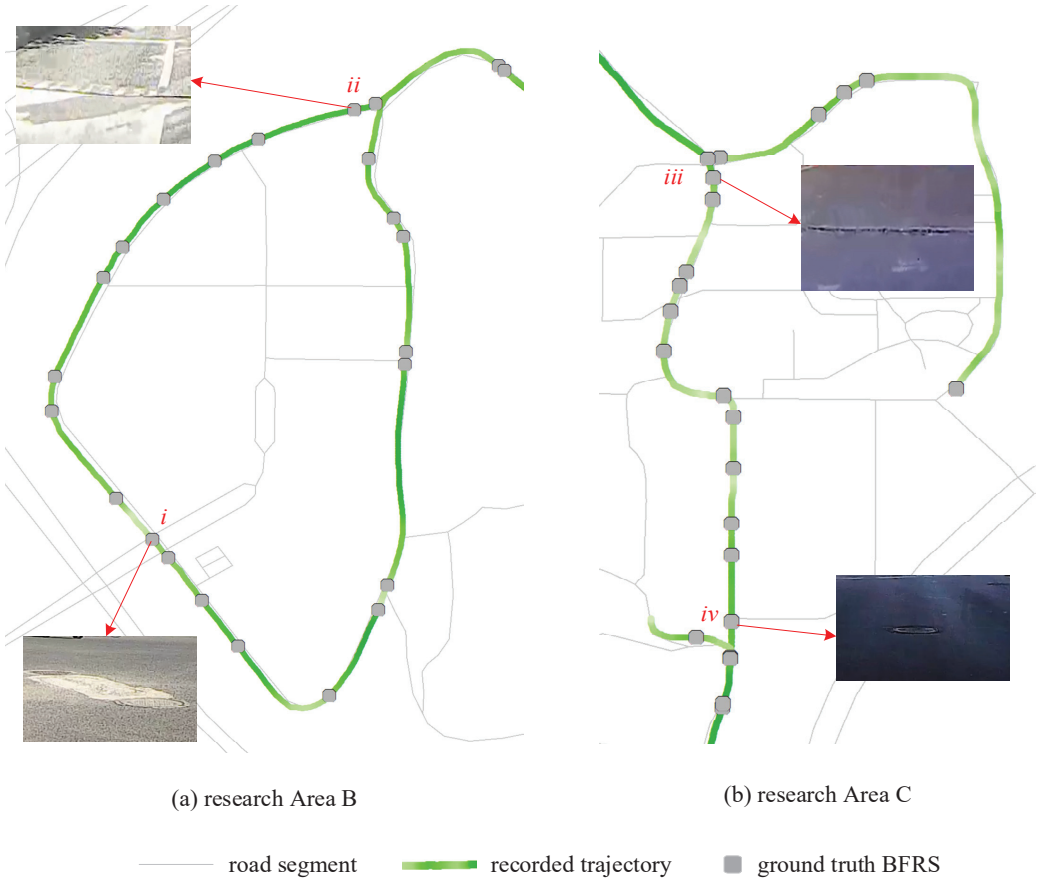


Figure 14. Different research areas. (**a**) Research Area B. (**b**) Research Area C.

In Figure 14, different kinds of BFRS in research Area B and C are visualized as the related pictures of (*i*) cracks, (*ii*) speed breaker, (*iii*) transverse gutter, and (*iv*) pothole. The detailed statistics of the collected multi-sensor stream data are depicted in Table 3.

Table 3. Statistic of collected multi-senor stream data in different areas.

	Length	Duration	GPS Records	Acceleration Records	Orientation Records
Area B	1171 m	234 s	234	23,233	23,233
Area C	1475 m	1298 s	1298	129,081	129,081

With the test dataset in Areas B and C, BFRS are firstly collected using the proposed method, refined and segmented as different BFRS groups, and clustered and represented as BFRS objects. Then, comparisons are conducted using the adaptive threshold value method, CWT method, and original CNN method, using the same parameter as that in Area A. Comparisons with different BFRS detection methods in Area B are depicted in Figure 15.

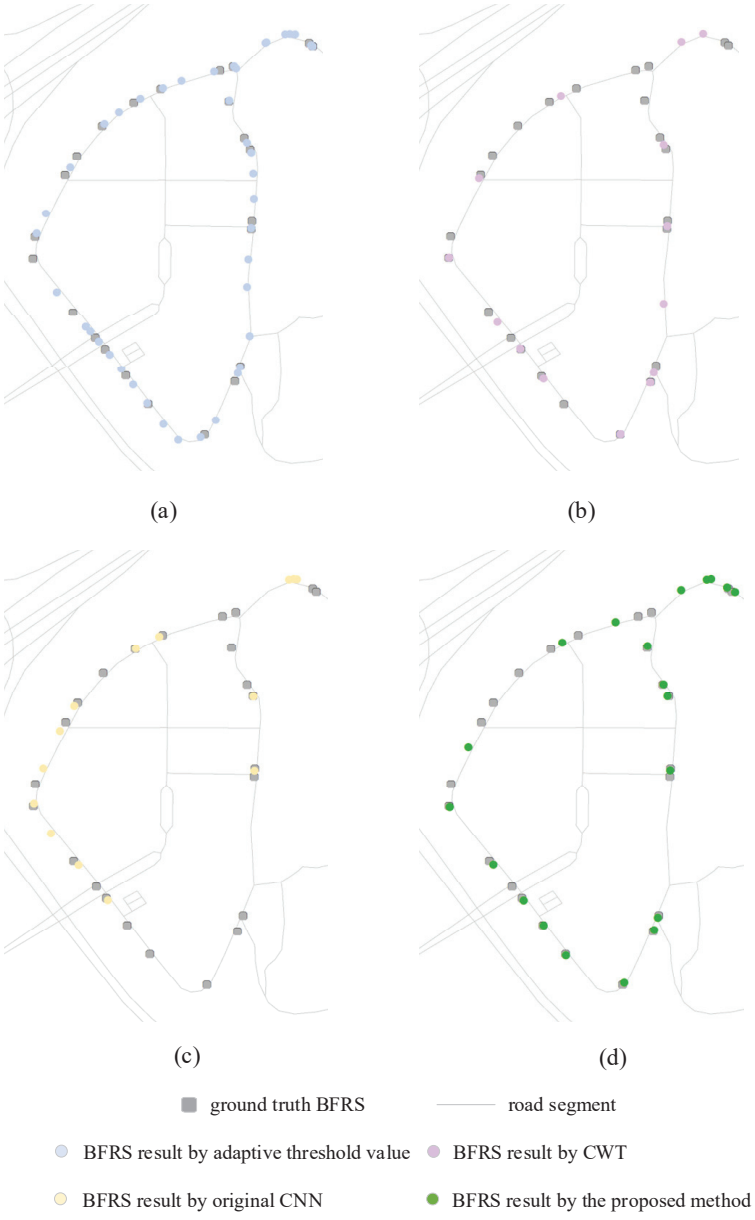


Figure 15. Comparisons with different BFRS detection methods in Area B. (**a**) Adaptive threshold value. (**b**) CWT. (**c**) Original CNN. (**d**) The proposed method.

There are 24 ground truth BFRS in Area B, while the number of detected BFRS varies in different results. In Figure 15a, there are 41 BFRS detected based on the adaptive threshold value method, and 18 BFRS are correctly detected. However, there are 23 wrong BFRS in the result, which holds the highest value compared with other methods. In Figure 15b,c, there are 14 BFRS detected in each result based on the CWT method and the original CNN method, while it performs a little better in the result of the CWT method with nine correct BFRS and five wrong BFRS. In contrast to the result of Figure 15c, the performance of the proposed method is dramatically improved, and 20 BFRS are detected, with 14 correct BFRS and 6 wrong BFRS, since the model is trained with the augmented dataset based on movement feature. The detailed statistics of the experiment results for Area B are depicted in Table 4.

Table 4. Statistics of the comparison results in Area B.

	Detected BFRS	**Correct BFRS**	**Wrong BFRS**	***Precision***	***Recall***	***F-Score***
Adaptive threshold value	41	18	23	43.90%	75.00%	0.5538
CWT	14	9	5	64.29%	37.50%	0.4737
Original CNN	14	8	6	57.14%	33.33%	0.4210
The proposed method	20	14	6	70.00%	58.33%	0.6363

In Table 4, the *precision*, *recall*, and *F-score* are computed according to the experiment result in Figure 15. Although there are 18 correct BFRS detected by the adaptive threshold value method, the number of wrong BFRS is also the highest among all of the results, and the related *TP, FP, and FN* are computed as 18, 23, and 6. Hence, the *precision*, *recall,* and *F-score* are 43.90%, 75.00%, and 0.5538 for the first method. With the similar detection result by the CWT method and the original CNN method, the related *TP*, *FP*, and *FN* are computed as CWT: {*TP*: 9, *FP*: 5, *FN*: 15} and original CNN: {*TP*: 8, *FP*: 6, *FN*: 16}, respectively. Moreover, the *precision*, *recall*, and *F-score* are 64.29%, 37.50%, and 0.4737 for the CWT method, and 57.14%, 33.33%, and 0.4210 for the original method. Finally, the *precision*, *recall*, *and F-score* for the proposed method {*TP*: 14, *FP*: 6, *FN*: 10} are 70.00%, 58.33%, and 0.6363. With the highest *precision*, *recall*, and *F-score*, the proposed method performs best among all results.

As shown in Figure 15 and Table 4 of Area B, where BFRS are more prevalent, detecting all BFRS is challenging. However, the adaptive threshold method detected a higher number of BFRS but resulted in many false detections, leading to low precision. The CWT method applies continuous wavelet transformation to the acceleration data, resulting in complex wavelet coefficient matrixes that are difficult to distinguish using specific thresholds. The original CNN method, lacking training on a dataset enhanced based on movement characteristics, did not perform well in the new experimental area, detecting only 14 BFRS. In contrast, the method proposed in this article, which is based on the analysis and training of vehicle motion characteristics, identified 20 BFRS with a high *precision* and the best relative *F-score*, demonstrating the superior performance of our method.

To further evaluate the performance of the proposed detection method, the BFRS detection operation is conducted in Area C, along with the related comparisons with the adaptive threshold value method, CWT method, and original CNN method, and uses the same parameter as that in Area A. The result is depicted in Figure 16.

In Area C, there are 23 ground truth BFRS spatially distributed, as shown by the gray rectangle in Figure 16. In Figure 16a, 27 BFRS are detected by the adaptive threshold value method, with 14 correct BFRS and 13 wrong BFRS. In Figure 16b,c, the detected BFRS are 11 by the CWT method and 9 by the original CNN method, and both of them perform similarly with 4 wrong BFRS; however, there are more correct BFRS in Figure 16b than that of Figure 16c. Compared with other results, there are 22 BFRS detected by the proposed method, and 17 BFRS are correctly detected, which is the highest value among all results. The detailed statistics of the experiment results for Area C are depicted in Table 5.

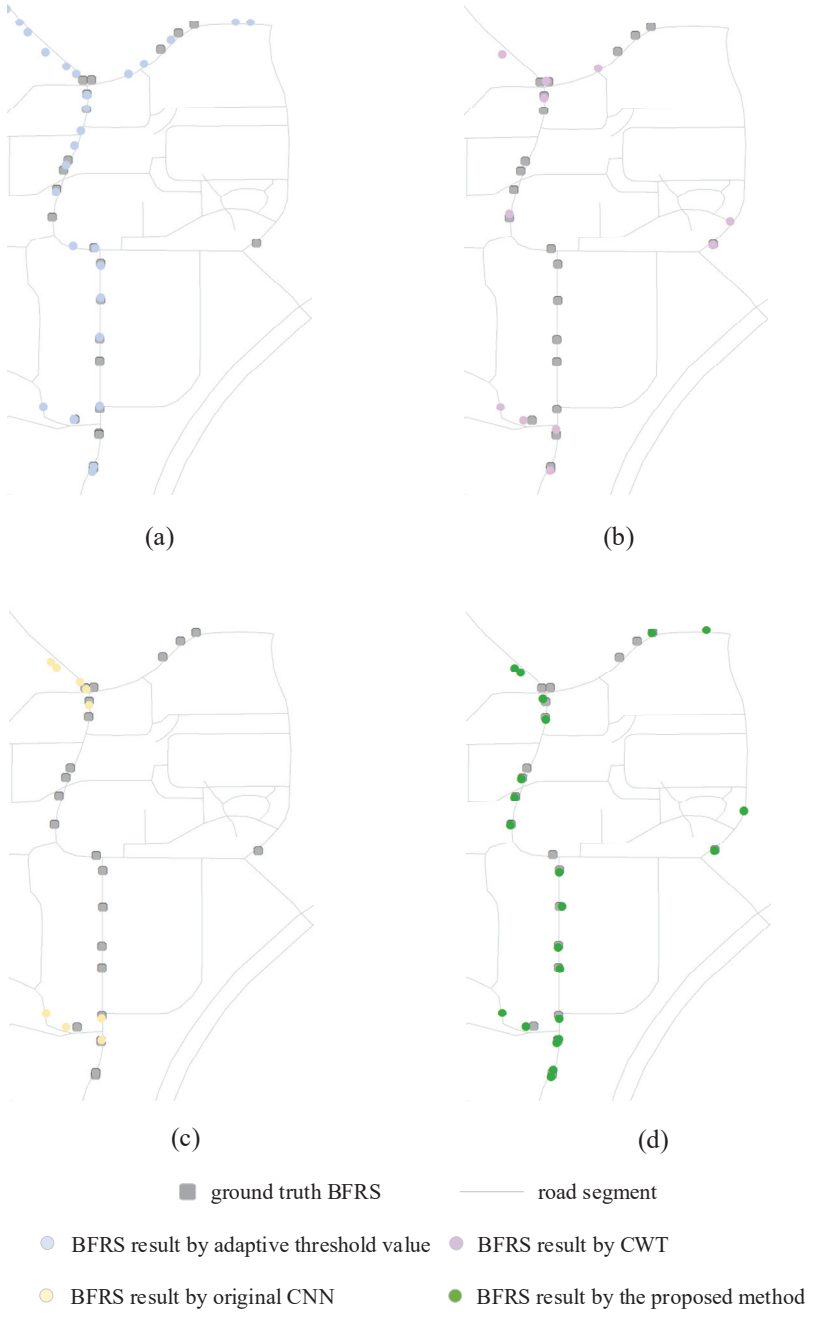


Figure 16. Comparisons with different BFRS detection methods in Area C. (**a**) Adaptive threshold value. (**b**) CWT. (**c**) Original CNN. (**d**) The proposed method.

Table 5. Statistics of the comparison results in Area C.

	Detected BFRS	Correct BFRS	Wrong BFRS	*Precision*	*Recall*	*F-Score*
Adaptive threshold value	27	14	13	51.85%	60.87%	0.56
CWT	11	7	4	63.64%	30.43%	0.4117
Original CNN	9	5	4	55.56%	21.74%	0.3125
The proposed method	22	17	5	77.27%	73.91%	0.7555

In Table 5, the *precision*, *recall*, and *F-score* are computed according to the experiment result in Figure 16. With 14 correct BFRS and 13 wrong BFRS by the adaptive threshold value method, the related *TP, FP, and FN* are computed as 14, 13, and 9, and the *precision*, *recall*,

and *F-score* are 60.87%, 75.00%, and 0.5600 for the result in Figure 16a. With seven correct BFRS and four wrong BFRS in Figure 16b by the CWT method, the result in Figure 16b {*TP*: 7, *FP*: 4, *FN*: 16} outperforms that of Figure 16c by the original CNN method {*TP*: 5, *FP*: 4, *FN*: 18}. However, the *precision and recall* of the proposed method are 77.27% and 73.91%, according to the statistic results {*TP*: 15, *FP*: 7, *FN*: 6}; hence, the *F-score* is the highest among all results {adaptive threshold value: 0.5600; CWT: 0.4117, original CNN: 0.3125; the proposed method: 0.7555}. More correct BFRS and less wrong BFRS are detected based on the proposed method, and it performs best among all experimented results.

As shown in Figure 16 and Table 5 of Area C, where the spatial distribution of BFRS is denser compared to the experimental Area B, the adaptive threshold method detected a larger number of BFRS, but many were incorrect, resulting in low *precision*. The CWT method showed higher detection *precision*, yet selecting appropriate BFRS based on wavelet coefficients remains challenging. The original CNN detected 9 BFRS, whereas the proposed method detected 22 BFRS, with 17 being correct BFRS. Therefore, compared to these methods, the *F-score* performance of the proposed method is the best, consistent with the results from Area B.

With experiment results in Figures 15 and 16, Tables 4 and 5, BFRS can be detected with high *precision*, *recall*, and *F-score* based on the proposed method. Although there are more BFRS detected by the adaptive threshold value method, the *precision* is not good, as there are also too many wrong BFRS in the result. The CWT method performs better than the adaptive threshold value with a higher *precision*, but the number of detected BFRS is small, as is the result of the original CNN method. However, using an augmented dataset based on the movement feature and the learnable parameters, BFRS are detected with the highest quality index, such as *precision*, *recall*, and *F-score*. Finally, the proposed method performs best in the comparison. The experiments demonstrate that the proposed method described in this article, which establishes a feature representation for BFRS through spectrum modeling, enhances the BFRS training dataset using vehicle motion characteristics, and detects BFRS in different experimental areas using the CNN, achieves relatively superior experimental results compared to other methods. This indicates the applicability and superior performance of the proposed method.

5. Conclusions and Future Work

Road infrastructure is a critical component of the geographic data required for the development of smart cities. It plays a pivotal role in city transportation systems, where the integrity of road surfaces directly influences transportation safety and the quality of the driving experience. The detection and management of road surface irregularities, such as BFRS, are essential due to the stresses imposed by extensive vehicle use and infrastructure wear. To detect BFRS in an economical way, smartphones assembled with multiple sensors are applied in this study. By harnessing the ubiquity and sophisticated sensor capabilities of these devices, discrete-sampled and non-homogeneous datasets are collected. The relationship between BFRS and vehicle speed conditions was analyzed, leading to the development of a robust model that effectively identifies BFRS. Through preprocessing of three-axis acceleration stream data, a comprehensive BFRS model is constructed and augmented by the training dataset with movement features. Then, the deep learning neural network is designed, and the multi-dimensional spectrum feature is input into the neural network. Finally, BFRS are refined and weighted clustered according to detected points and the related probability. Multi-sensor stream data at different areas are collected, training datasets are augmented, spectrum features are constructed based on the modeling result, and the deep learning neural network is trained and validated. The proposed approach not only demonstrated superior performance in detecting BFRS compared to other existing methods, but also provided initial locations of the distress or damages of the road surface, which would contribute to the maintenance of road assets. The flexibility of the CNN model, which can be adapted or replaced with alternative architectures, adds to the scalability of the approach. The smartphone sensors used in this article are highly integrated within

the device. However, in IoT sensor application scenarios, especially when dealing with multiple sensors that lack necessary external enclosures and have lower levels of internal integration, calibrating sensors is essential. Additionally, the external environment (such as external magnetic fields) and the impact of different disturbances require in-depth study.

Future research aims to improve the performance of BFRS detection results with multi-modal sensors; for example, since video datasets can also be collected using smartphones, the deep learning neural network framework can be improved and redesigned to encode multi-modal stream data. Moreover, for vehicles with different types of tires and suspension systems, it is essential to explore the analysis of forces while overcoming an obstacle. This could provide deeper insights into the interaction between the vehicle's tires, suspension, and road surface, which could be crucial for encoding the BFRS and interpreting the results more accurately. Another aspect that needs to be further explored is BFRS detection with crowdsourced datasets, which collect road surface conditions using volunteered geographic information.

Author Contributions: H.L. was responsible for methodology, analysis, and drafting of the manuscript. Q.Z. contributed to the implementation, programming, and testing. D.J. was responsible for literature reviews, background knowledge, and improvement of the writing. J.H. contributed to the data collection and visualization of the results. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China [Grant No. 42101466].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that were analyzed in this study can be found at https://doi.org/10.6084/m9.figshare.26039233.v1. The street map is available from OpenStreetMap, accessed on 19 March 2024.

Acknowledgments: We would like to give special thanks to Dieter Pfoser at George Mason University for their comments and revisions that have improved the quality of this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Li, W.; Batty, M.; Goodchild, M.F. Real-time GIS for smart cities. *Int. J. Geogr. Inf. Sci.* **2020**, *34*, 311–324. [CrossRef]
2. Huang, H.; Yao, X.A.; Krisp, J.M.; Jiang, B. Analytics of location-based big data for smart cities: Opportunities, challenges, and future directions. *Comput. Environ. Urban Syst.* **2021**, *90*, 101712. [CrossRef]
3. Li, X.; Goldberg, D.W. Toward a mobile crowdsensing system for road surface assessment. *Comput. Environ. Urban Syst.* **2018**, *69*, 51–62. [CrossRef]
4. Li, D.; Shan, J.; Shao, Z.; Zhou, X.; Yao, Y. Geomatics for smart cities-concept, key techniques, and applications. *Geo-Spat. Inf. Sci.* **2013**, *16*, 13–24. [CrossRef]
5. Wei, Y.; Zhang, K.; Ji, S. Simultaneous road surface and centerline extraction from large-scale remote sensing images using CNN-based segmentation and tracing. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 8919–8931. [CrossRef]
6. Yi, C.W.; Chuang, Y.T.; Nian, C.S. Toward crowdsourcing-based road pavement monitoring by mobile sensing technologies. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 1905–1917. [CrossRef]
7. Li, X.; Huo, D.; Goldberg, D.W.; Chu, T.; Yin, Z.; Hammond, T. Embracing crowdsensing: An enhanced mobile sensing solution for road anomaly detection. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 412. [CrossRef]
8. Rajamohan, D.; Gannu, B.; Rajan, K.S. MAARGHA: A prototype system for road condition and surface type estimation by fusing multi-sensor data. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 1225–1245. [CrossRef]
9. Dong, D.; Li, Z. Smartphone sensing of road surface condition and defect detection. *Sensors* **2021**, *21*, 5433. [CrossRef]
10. Wang, S.; Yang, F.; Cheng, Y.; Yang, Y.; Wang, Y. Adaboost-based crack detection method for pavement. In *IOP Conference Series: Earth and Environmental Science*; IOP Publishing: Bristol, UK, 2018; Volume 189, p. 022005. [CrossRef]
11. Ma, N.; Fan, J.; Wang, W.; Wu, J.; Jiang, Y.; Xie, L.; Fan, R. Computer vision for road imaging and pothole detection: A state-of-the-art review of systems and algorithms. *Transp. Saf. Environ.* **2022**, *4*, tdac026. [CrossRef]

12. Jenkins, M.D.; Carr, T.A.; Iglesias, M.I.; Buggy, T.; Morison, G. A deep convolutional neural network for semantic pixel-wise segmentation of road and pavement surface cracks. In Proceedings of the 2018 26th European Signal Processing Conference (EUSIPCO), Rome, Italy, 3–7 September 2018; pp. 2120–2124. [CrossRef]
13. Zhang, Z.; Ai, X.; Chan, C.K.; Dahnoun, N. An efficient algorithm for pothole detection using stereo vision. In Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014; pp. 564–568. [CrossRef]
14. Xiao, L.; Dai, B.; Liu, D.; Hu, T.; Wu, T. CRF based road detection with multi-sensor fusion. In Proceedings of the 2015 IEEE Intelligent Vehicles Symposium (IV), Seoul, Republic of Korea, 28 June–1 July 2015; pp. 192–198. [CrossRef]
15. Fan, R.; Ozgunalp, U.; Hosking, B.; Liu, M.; Pitas, I. Pothole detection based on disparity transformation and road surface modeling. *IEEE Trans. Image Process.* **2019**, *29*, 897–908. [CrossRef] [PubMed]
16. Kulambayev, B.; Beissenova, G.; Katayev, N.; Abduraimova, B.; Zhaidakbayeva, L.; Sarbassova, A.; Akhmetova, O.; Issayev, S.; Suleimenova, L.; Kasenov, S.; et al. A Deep Learning-Based Approach for Road Surface Damage Detection. *Comput. Mater. Contin.* **2022**, *73*, 3403–3418. [CrossRef]
17. Chen, C.; Seo, H.; Zhao, Y. A novel pavement transverse cracks detection model using WT-CNN and STFT-CNN for smartphone data analysis. *Int. J. Pavement Eng.* **2022**, *23*, 4372–4384. [CrossRef]
18. Siegemund, J.; Franke, U.; Förstner, W. A temporal filter approach for detection and reconstruction of curbs and road surfaces based on conditional random fields. In Proceedings of the 2011 IEEE Intelligent Vehicles Symposium (IV), Baden-Baden, Germany, 5–9 June 2011; pp. 637–642. [CrossRef]
19. Kang, B.; Choi, S. Pothole detection system using 2D LiDAR and camera. In Proceedings of the 2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN), Milan, Italy, 4–7 July 2017; pp. 744–746. [CrossRef]
20. Bhatt, A.; Bharadwaj, S.; Sharma, V.B.; Dubey, R.; Biswas, S. An Overview of Road Health Monitoring System for Rigid Pavement By Terrestrial Laser Scanner. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2022**, *43*, 173–180. [CrossRef]
21. Harikrishnan, P.M.; Gopi, V.P. Vehicle vibration signal processing for road surface monitoring. *IEEE Sens. J.* **2017**, *17*, 5192–5197. [CrossRef]
22. Yuan, Y.; Islam, S.; Yuan, Y.; Wang, S.; Baker, T.; Kolbe, L.M. EcRD: Edge-cloud computing framework for smart road damage detection and warning. *IEEE Internet Things J.* **2020**, *8*, 12734–12747. [CrossRef]
23. Singh, G.; Bansal, D.; Sofat, S.; Aggarwal, N. Smart patrolling: An efficient road surface monitoring using smartphone sensors and crowdsourcing. *Pervasive Mob. Comput.* **2017**, *40*, 71–88. [CrossRef]
24. Mednis, A.; Strazdins, G.; Zviedris, R.; Kanonirs, G.; Selavo, L. Real time pothole detection using android smartphones with accelerometers. In Proceedings of the 2011 International Conference on Distributed Computing in Sensor Systems and Workshops (DCOSS), Barcelona, Spain, 27–29 June 2011; pp. 1–6. [CrossRef]
25. Nunes, D.E.; Mota, V.F.S. A participatory sensing framework to classify road surface quality. *J. Internet Serv. Appl.* **2019**, *10*, 13. [CrossRef]
26. Xu, C.; Zhang, Q.; Mei, L.; Chang, X.; Ye, Z.; Wang, J.; Ye, L.; Yang, W. Cross-Attention-Guided Feature Alignment Network for Road Crack Detection. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 382. [CrossRef]
27. Li, L.; Sun, L.; Ning, G.; Tan, S. Automatic pavement crack recognition based on BP neural network. *Promet-Traffic Transp.* **2014**, *26*, 11–22. [CrossRef]
28. Azhar, K.; Murtaza, F.; Yousaf, M.H.; Habib, H.A. Computer vision based detection and localization of potholes in asphalt pavement images. In Proceedings of the 2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), Vancouver, BC, Canada, 15–18 May 2016; pp. 1–5. [CrossRef]
29. Ouma, Y.O.; Hahn, M. Pothole detection on asphalt pavements from 2D-colour pothole images using fuzzy c-means clustering and morphological reconstruction. *Autom. Constr.* **2017**, *83*, 196–211. [CrossRef]
30. Sharma, S.K.; Phan, H.; Lee, J. An application study on road surface monitoring using DTW based image processing and ultrasonic sensors. *Appl. Sci.* **2020**, *10*, 4490. [CrossRef]
31. Mednis, A.; Strazdins, G.; Liepins, M.; Gordjusins, A.; Selavo, L. RoadMic: Road surface monitoring using vehicular sensor networks with microphones. In Proceedings of the Networked Digital Technologies: Second International Conference, NDT 2010, Prague, Czech Republic, 7–9 July 2010; Proceedings, Part II 2. Springer: Berlin/Heidelberg, Germany, 2010; pp. 417–429. [CrossRef]
32. Allouch, A.; Koubaa, A.; Abbes, T.; Ammar, A. Roadsense: Smartphone application to estimate road conditions using accelerometer and gyroscope. *IEEE Sens. J.* **2017**, *17*, 4231–4238. [CrossRef]
33. Mihoub, A.; Krichen, M.; Alswailim, M.; Mahfoudhi, S.; Salah, R.B.H. Road scanner: A road state scanning approach based on machine learning techniques. *Appl. Sci.* **2023**, *13*, 683. [CrossRef]
34. Zang, K.; Shen, J.; Huang, H.; Wan, M.; Shi, J. Assessing and mapping of road surface roughness based on GPS and accelerometer sensors on bicycle-mounted smartphones. *Sensors* **2018**, *18*, 914. [CrossRef] [PubMed]
35. Basavaraju, A.; Du, J.; Zhou, F.; Ji, J. A machine learning approach to road surface anomaly assessment using smartphone sensors. *IEEE Sens. J.* **2019**, *20*, 2635–2647. [CrossRef]
36. Varona, B.; Monteserin, A.; Teyseyre, A. A deep learning approach to automatic road surface monitoring and pothole detection. *Pers. Ubiquitous Comput.* **2020**, *24*, 519–534. [CrossRef]

37. Ren, M.; Zhang, X.; Chen, X.; Zhou, B.; Feng, Z. YOLOv5s-M: A deep learning network model for road pavement damage detection from urban street-view imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *120*, 103335. [CrossRef]
38. Pachori, R.B. *Time-Frequency Analysis Techniques and Their Applications*; CRC Press: Boca Raton, FL, USA, 2023. [CrossRef]
39. Kim, D.-Y.; Lee, H.W.; Yu, J.; Park, J.-K. Application of Convolutional Neural Networks for Classifying Penetration Conditions in GMAW Processes Using STFT of Welding Data. *Appl. Sci.* **2024**, *14*, 4883. [CrossRef]
40. Fukunaga, K.; Larry, H. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Trans. Inf. Theory* **1975**, *21*, 32–40. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

sustainability

MDPI

Article

Smart-City Policy in China: Opportunities for Innovation and Challenges to Sustainable Development

Song Yang [1,2,*], Yinfeng Su [1] and Qin Yu [3]

[1] School of Mathematics and Statistics, Guizhou University of Finance and Economics, Guiyang 550025, China
[2] Guizhou Key Laboratory of Big Data Statistical Analysis, Guiyang 550025, China
[3] School of Accounting, Nanjing University of Finance and Economics, Nanjing 210000, China
* Correspondence: song_yang@mail.gufe.edu.cn

Abstract: Urban development relies on the promotion of innovation, while sustainable development is an inevitable requirement for green urban development. As the primary carrier of innovation and sustainable development, cities have seen the construction of smart cities become a hotspot topic of public concern against the backdrop of rapid advancements in information technology. Based on the Chinese smart-city pilot policies, this paper selects data from 278 prefecture-level cities between 2007 and 2020, employing difference-in-difference (DID), epsilon-based measures and global Malmquist–Luenberger index (EBM-GLM), and the Spatial Durbin Model (SDM) to analyze the direct impact, spatial effects, and regional differences of smart-city construction on urban innovation capacity and sustainable development. The research results indicate the following: (1) the implementation of smart-city policies significantly enhances the urban innovation capacity (*UCI*), but its impact on green total-factor productivity (*GTFP*) is unstable and even insignificant; (2) the *UCI* and *GTFP* of smart cities have spillover effects, and the implementation of policies may inhibit the improvement of *UCI* and *GTFP* in neighboring cities; (3) the impact of smart-city construction varies across different regions, with a more significant promotion effect on the innovation capacity of economically developed cities.

Keywords: smart city; innovation capacity; sustainable development; regional difference; EBM-GLM; SDM; DID

1. Introduction

The advent of the information age has profoundly changed people's lifestyles. The emergence of smart products and systems, such as smart furniture, intelligent transportation, intelligent financial management, and smart supervision, has brought great convenience to people's lives. Some scholars argue that knowledge and the environment have geographical "stickiness", and innovation and sustainable development also exhibit regional agglomeration effects. This phenomenon is particularly evident at the local city level, thus more effectively promoting urban intelligentization [1]. Therefore, as important carriers of innovation and sustainable development, cities should fully leverage their roles. By strengthening planning, optimizing industrial structures, and improving environmental quality, cities can be developed into livable and workable green spaces, thereby promoting high-quality economic development.

The development of smart cities can be traced back to the mid-1970s. At that time, Los Angeles launched the first large-scale urban data project [2]. Subsequently, various countries formulated smart-city construction policies in the urban dimension to promote the development of smart cities, enhance the convenience of residents' lives, and improve the living environment. Examples include Barcelona's Smart City initiative [3], smart cities distributed across North America, South America, Europe, Africa, Oceania, and Asia [4–6], smart-city politics in Italy [7], and Australian smart cities [8]. In order to transform the

traditional extensive urban development model, the Chinese government launched the National Smart City Policy (SCP) in 2012, which is based on pilot cities, integrating concepts such as "intensive", "low-carbon", "ecological", and "smart" into urban environmental construction. This initiative aims to shape the urban landscape of the Innovation 2.0 era and lead to a city with intelligence bearing Chinese characteristics. Therefore, what is a smart city? What impact does the implementation of smart-city policies have on urban development? These have become hot topics in smart-city research.

Regarding the definition of smart cities, some scholars argue that smart cities are not cities with specific attributes but, rather, a characterization of different contexts within an urban space [**?**]. Other scholars believe smart cities primarily refer to cities that integrate infrastructure and technological-intermediary services, enhance social learning through human infrastructure, and merge governance through institutional and citizen participation across three dimensions [4]. Concerning the impact of smart-city construction on urban development, existing research has found that smart cities can utilize technology and data to improve the efficiency, economic development, sustainability, and quality of life of urban residents [10]. Additionally, some scholars have utilized the directional distance function (DDF) and Luenberger productivity indicators to measure the green total-factor productivity ($GTFP$) of cities, analyzing the causal relationship between smart-city policies (SCP) and $GTFP$. The results indicate that SCP significantly promotes the $GTFP$ of Chinese cities, mainly through technological innovation, industrial structure upgrading, and resource allocation optimization [11]. Stübinger and Schneider (2020) [**?**] reviewed the top 200 related research papers ranked by Google Scholar, concluding that intelligence can minimize energy, water, food, waste, heat output, and air pollution and affirming the importance of intelligent, sustainable development in the future. However, most existing studies have focused on the impact of smart-city construction on urban technology, carbon emissions, and $GTFP$ based on DDF–Luenberger. Urban innovation capability is a critical indicator of a city's future development potential, and green urban development is an essential aspect of sustainable urban development. Currently, there is a lack of research on comprehensive urban innovation capabilities and $GTFP$ based on EBM-GLM. Therefore, this paper systematically explores the impact of smart-city policy implementation on the innovation capabilities and green total-factor productivity of Chinese cities while investigating spatial effects and regional differences.

The marginal contributions of this research are as follows:

- based on the measurement standards for the urban innovation capability index provided in the report "China's City and Industry Innovation Report 2017" jointly released by the First Financial Research Institute and Fudan University [13] (http://www.360doc.com/content/18/0505/21/26988834_751429774.shtml (accessed on 28 June 2024)), this paper constructs the Urban Innovation Capability Index from two micro-level data: an innovation index (patent maintenance) and an entrepreneurship index (the establishment of enterprises). It then employs the DID and SDM-DID models [14] to analyze both the direct and spatial indirect effects of smart-city construction on urban innovation capability in China.
- Secondly, this paper uses EBM to calculate the city's green total-factor productivity ($GTFP$). It decomposes $GTFP$ into two parts using GML [15]: the green efficiency change (GEC) index and the green technical change (GTC) index. From the perspective of urban network externalities, it reassesses the spatial impact of smart cities on urban green total-factor productivity, urban green efficiency change, and urban technological progress using the DID and SDM-DID models, thereby supplementing the academic discussion on issues related to smart-city network externalities.
- Thirdly, this paper categorizes cities based on geographic regions and economic development levels in China, conducts regression on subsamples, and analyzes the regional differences in the impact of smart-city policy implementation on urban innovation capability and urban green total-factor productivity, thus expanding the research perspective on smart cities.

2. Literature Review

With the continuous advancement of information technology, smart cities increasingly represent a significant result of the deep integration of urban modernization and informatization. The existing research has primarily focused on three aspects: the definition of smart cities, the factors influencing the construction of smart cities, and the impact of smart-city construction on urban innovation and urban sustainable development.

2.1. Related Research on the Definition of a Smart City

In the study of smart cities, scholars have engaged in extensive discussions on the question of "what a smart city is". However, the definitions of a smart city are varied, and there is currently no unified definition. Nam and Pardo (2011) [4] compiled various interpretations. Hollands (2008) [16] views the smart city as a "city label" phenomenon, particularly in terms of what the label reveals and conceals ideologically. Washburn (2010) [17] defined it as the use of smart computing technologies to make the critical infrastructure components and services of a city—which include city administration, education, healthcare, public safety, real estate, transportation, and utilities—more intelligent, interconnected, and efficient. NRDC (2008) [18] sees a city striving to make itself "smarter" (more efficient, sustainable, equitable, and livable), proposing a definition of "A city that gives inspiration, shares culture, knowledge, and life, a city that motivates its inhabitants to create and flourish in their own lives". Marsa-Maestre et al. (2008) [19] also believe that smart means more user-friendly than intelligent, as the latter is limited to quick thinking and responding to feedback. A smart city needs to adapt to user needs and provide a customized interface. Therefore, the development of smart cities is multi-faceted. Initially, scholars applied the name to various small dimensions. For example, a digital city [20], an intelligent city [21], an information city [22], a creative city [23], a learning city [24], and so on.

With the advancement of technology, people's understanding of smart cities has become increasingly comprehensive, and scholars have begun to analyze this concept from more dimensions. Dameri et al. (2013) [25] believe the definition of a smart city is not meant to address a theoretical need but to support the public governance of smart cities by defining visions, goals, and policies that can promote concrete smart implementations. Kim et al. (2017) [26] argue that the main characteristics of a smart city are the high integration of information technology and the comprehensive application of information resources. The fundamental elements of smart-city development include smart technology, smart industry, smart services, smart management, and smart living. Kirimtat et al. (2020) [27] analyzed smart systems from the perspectives of smart people, a smart economy, smart governance, smart mobility, a smart environment, and smart living. Yigitcanlar et al. (2021) [8] argue that innovation, sustainability, and governance are the most popular concepts of smart cities, and the Internet of Things, artificial intelligence, and autonomous vehicle technologies are the most popular technologies.

2.2. Related Research on Factors Influencing Smart-City Construction

Some scholars primarily explore the factors affecting the construction of smart cities from both macro and micro perspectives. On the macro level, urban environmental factors, economic development, structural urban variables, infrastructure, smart-city services, the Internet of Things (IoT), and modern government services are often considered vital factors influencing smart-city development. Neirott et al. (2014) [28] argue that the evolutionary patterns of smart cities (SCs) are highly dependent on local environmental factors. Specifically, economic development and structural urban variables may influence a city's digital trajectory, geographical location may shape SC strategies, and population density and its associated congestion issues may be crucial components in determining SC implementation pathways. Kim (2022) [29], on the other hand, believes that the construction of urban infrastructure and smart-city services are key factors influencing smart-city development. Rejeb et al. (2022) [30] analyze the main application scenarios of the IoT in smart cities from an IoT perspective, including smart buildings, transportation, healthcare,

smart parking, and smart grids. Yaqoob et al. (2023) [31] emphasize the promotion of smart-city construction through factors such as improving infrastructure, modernizing government services, increasing accessibility, accelerating economic growth, and promoting sustainability. Additionally, Alizadeh and Sharifi (2023) [32] consider social justice a critical aspect of smart-city construction. Kruhlov et al. (2024) [33] view the construction of PPP projects as an effective means to promote urban development based on the concept of smart cities.

On the micro level, the analysis focuses on clean technology, information and communication technology (ICT), and governance mechanisms. Lai et al. (2020) [10] believe that clean technology can advance the development of smart cities, including energy, transportation, and sanitation. Chen (2021) [34] states that most components of a smart city rely heavily on ICT. Mora et al. (2023) [35] argue that governance mechanisms are crucial for the transformation of smart cities. However, some scholars contend that many studies have emphasized the benefits of smart cities while neglecting the drawbacks of technology and failed projects, as well as the potential impacts on data privacy and security, which should be given more attention [36,37].

2.3. The Impact of Smart-City Construction on Urban Innovation and Sustainable Development

Cities are the primary carriers of innovation, and sustainable development constitutes a significant component of green cities. Therefore, the existing research has primarily analyzed the impact of smart-city construction on urban innovation and urban sustainable development.

In terms of urban innovation, Bibri and Krogstie (2020) [38] found, by comparing London and Barcelona, two leading data-driven smart cities in Europe, that these cities have a high level of development in applied data-driven technologies. It can thus be inferred that smart-city construction can drive the growth of applied data-driven technologies. Yang et al. (2024) [39], based on micro-enterprise data from the China Taxation Survey Database and the China Innovation Enterprise Database for the period of 2010–2015, found that China's Smart City Pilot Policy (SCPP) can encourage firm-level green innovation. Additionally, the impact of green innovation on SCPP is more pronounced for non-state-owned enterprises, heavily polluting industries, and non-resource-based city enterprises compared to state-owned enterprises, lightly polluting industries, and resource-based city enterprises. Gohar and Nencioni (2021) [40] and Alahi et al. (2023) [41] argue that smart cities rely on information and communication technology (ICT), and the construction of smart cities can also stimulate the development of urban ICT, such as the construction of 5G communication infrastructure. Kashef et al. (2021) [42] analyzed the demand and impact of smart-city construction on the enhancement of wireless integrated mesh technology (WIMTE) smart-city monitoring systems. Furthermore, scholars have found that smart-city construction helps improve urban innovation performance, although this impact may vary between cities, with each city demonstrating some unique results [43]. Concurrently, the effect of smart-city construction on urban innovation mainly manifests in an increase in innovation activities [44], the enhancement of informatization levels, the strengthening of governmental support for science and technology, the optimization of industrial structures [45], the promotion of green patents [46], green technological innovation, and enterprise innovation [47].

In the field of urban sustainable development, Chu et al. (2021) [48] explored the impact of smart-city construction on the quality of the ecological environment in China. They found that, from 2005 to 2017, China's smart-city initiatives reduced industrial waste gas and industrial wastewater by approximately 20.7% and 12.2%, respectively. Guo et al. (2022) [43], focusing on China's smart-city pilot policies, discovered that smart-city construction significantly reduced per-capita CO_2 emissions, with a reduction effect of about 18.42 log points. Moreover, the impact of smart-city construction on CO_2 emission reduction is more notable in cities with higher administrative levels, higher levels of neutral technological progress and green innovation, and more advanced industrial structures.

Meanwhile, Hoang et al. (2021) [49] and Clement et al. (2023) [50] posit that the utilization of renewable energy can reduce pollutant emissions and improve living environment quality. Integrating renewable energy into the energy system of smart cities could achieve more sustainable development. Chen et al. (2022) [51] suggest that energy-saving intelligent street lighting can save substantial energy during peak and off-peak periods, ultimately reducing energy consumption and carbon emissions. Meanwhile, Salman and Hasar (2023) [52] believe that intelligent traffic monitoring and management can reduce urban air pollution.

In addition, Wang et al. (2022) [11] found that SCP policies have a significant positive impact on the green total-factor productivity (*GTFP*) of Chinese cities and exhibit positive spatial spillover effects on the *GTFP* of adjacent non-pilot cities. However, the spillover effects on neighboring pilot cities are poorer. Nonetheless, Yang et al. (2024) [39] indicate that, in the short term, green innovation in China may exert some adverse effects on green total-factor productivity, but its long-term effects are positive. Additionally, some scholars have found that smart-city construction helps promote innovative urban development and enhance sustainable development and living quality [53], propelling governmental digital governance capabilities and thereby advancing green sustainable development [54].

In addition to researching the impact of smart-city construction on urban innovation and sustainable development, there have also been discussions on other aspects. Sharifi et al. (2021) [55] claim that smart-city initiatives can strengthen resilience against the COVID-19 pandemic and similar future events. Lee et al. (2023) [56] conducted surveys in Amsterdam (the Netherlands), Seoul (South Korea), Portland (USA), and Ho Chi Minh City (Vietnam), concluding that smart-city construction can promote the equity of access to innovative technologies, addressing issues of unfair access and mistrust. Clement et al. (2023) [57] highlight the importance of smart-city strategies in enhancing collaboration between local governments and stakeholders within an ecosystem.

From the above research, it can be seen that, although scholars have analyzed the impact of smart-city construction on urban green total-factor productivity (*GTFP*), their conclusions are inconsistent [11,39]. Additionally, Wang K-L. et al. (2022) [11] primarily used the Luenberger productivity index derived from the directional distance function (DDF) to measure urban *GTFP*. Furthermore, the existing research on the impact of smart cities on urban innovation has mainly been discussed from the perspective of technological innovation, with relatively scant discussion on comprehensive urban innovation capabilities. Therefore, this paper was intended to construct a comprehensive urban innovation capability index based on the microdata of the innovation index and the entrepreneurship index, according to the calculation standards of urban innovation capability in the "China Urban and Industrial Innovation Report 2017" [13]. It employs the EBM-GML method to measure urban *GTFP*, the green efficiency change (*GEC*) index, and the green technical change (*GTC*) index. By adopting a different measurement perspective, this study discusses the impact of smart cities on urban innovation and sustainable development, which can validate existing research conclusions and enrich the analysis of the effects of smart cities.

3. Research Hypotheses, Methods, and Models

3.1. Research Hypotheses

Innovation is the driving force behind urban development, and the city serves as the primary carrier of innovation. The enhancement of a city's innovation capacity is mainly reflected in two aspects. On the one hand, there is urban infrastructure construction. Existing research indicates that the development of smart cities relies primarily on information and communication technology (ICT). The construction of smart cities can also drive the growth of urban ICT, such as the establishment of 5G communication infrastructure [40,41]. On the other hand, there are advancements and innovations in core urban technologies. Studies have found that smart-city construction can propel the development of data-driven technologies [38], enhance the construction of urban wireless integrated mesh technology (WIMTE) smart-city monitoring systems [42], increase innovative activities [44], improve

informatization, strengthen government support for science and technology, and optimize industrial structures [45,58]. It also promotes the increase in green patents [46], green technology innovation, and corporate innovation [47], thereby enhancing urban innovation performance [43].

Moreover, Yang et al. (2024) [39], based on micro-enterprise data from the Chinese Tax Survey Database and the China Innovation Enterprise Database from 2010 to 2015, found that China's Smart City Pilot Policy (SCPP) can encourage green innovation at the enterprise level. Additionally, some researchers believe smart cities, by utilizing advanced information and communication technologies, have achieved intelligent and refined urban management [59,60]. Furthermore, the data sharing and openness of smart cities also promote the aggregation and sharing of innovation resources, improving the efficiency of innovation resource utilization [61]. This helps enhance urban operational efficiency, optimize resource allocation, and create a more favorable environment for innovative activities. By introducing the next generation of information technologies, smart cities promote a deep integration of traditional industries with information technology, giving rise to new industries, new business forms, and new models. This not only provides a new impetus for urban innovation but also promotes the sustainable development of urban economies. Therefore, this paper's first null hypothesis (H1_0) and alternative hypothesis (H1_1) are stated as follows:

Null hypothesis (H1_0). *The implementation of smart-city policies cannot significantly improve urban innovation capabilities.*

Alternative hypothesis (H1_1). *The implementation of smart-city policies can improve urban innovation capabilities.*

The construction of smart cities, in addition to influencing urban innovation, has been shown to exert a significant impact on urban sustainable development. This impact is analyzed from both positive and negative perspectives.

Regarding the positive impacts, research has found that smart-city development can reduce the emissions of industrial waste gas and wastewater [48], lower per-capita carbon dioxide emissions [43], increase the use of renewable energy, and reduce pollutant emissions [49,51,57], ultimately improving the quality of the ecological environment. Salman and Hasar (2023) [52] also found that intelligent traffic monitoring and management can reduce urban air pollution. Additionally, researchers have discovered that SCP policies have a significantly positive impact on the green total-factor productivity ($GTFP$) of cities [11]. Smart city development helps promote urban innovation, enhance sustainable development, and improve the quality of life [53]. It also advances governmental digital governance capabilities, thereby promoting green and sustainable development [54].

As for the negative impacts, Yang et al. (2024) [39] indicated that China's green innovation might have a particular adverse effect on green total-factor productivity in the short term. Simultaneously, the pursuit of innovative development in smart cities might sometimes suppress urban sustainable development due to the neglect of specific issues. These issues include inappropriate technology applications, the energy consumption of data centers, environmental neglect, and data security concerns. In the construction of smart cities, over-reliance on high-energy-consumption technologies or the introduction of technology without fully considering its environmental impact could lead to increased energy consumption and environmental pollution, thus hindering green urban development [62]. Smart cities usually involve extensive data collection, storage, and processing, which typically necessitates the construction of large data centers. Data centers are highly energy-consuming facilities, and if not built and managed correctly, they could become a bottleneck for green urban development [63]. Additionally, in the quest for innovation, if the construction of smart cities neglects socio-environmental impact assessments, it might lead to the implementation of projects or policies detrimental to sustainable development [64].

Furthermore, poor management, a lack of attention to information security, and technological leaks can lead to data security issues such as identity verification, unauthorized access, device-level vulnerabilities, and sustainability [65], resulting in social problems like extremism, polarization, misinformation, and internet addiction [66]. Ismagilova et al. (2020) [67] and Laufs et al. (2020) [68] also asserted that increasing research is focusing on the security, privacy, and risks associated with smart cities, highlighting the threats related to information security and the challenges in managing and processing personal data within smart-city infrastructures. Spicer et al. (2023) [69] used unique survey data from residents of three prominent Canadian cities—Vancouver, Montreal, and Toronto—to study inconsistency between the types of projects pursued by cities, in particular service and policy areas, and the preferences of the residents. However, the calculation of green total-factor productivity mainly includes the input and output sub-components through factoring in energy consumption and environmental issues within the traditional total-factor productivity analysis framework, reflecting the coordination between the economy and the environment. Therefore, combining the above analyses, the impact of smart cities on urban green total-factor productivity might be both positive and negative. Accordingly, this paper's second null hypothesis (H2_0) and alternative hypothesis (H2_1) are stated as follows:

Null hypothesis (H2_0). *The impact of smart city policies on urban green total-factor productivity is unstable and may even be insignificant.*

Alternative hypothesis (H2_1). *The impact of smart city policies on urban green total-factor productivity is significantly positive or negative.*

The development of cities often exhibits certain network externalities that can have an impact on the development of neighboring cities. Wang et al. (2022) [11] found that SCP had a positive spatial spillover effect on the *GTFP* of adjacent non-pilot cities, but the spillover effect on adjoining pilot cities was relatively poor. In addition, some studies have found that smart cities provide innovative talents with more convenient work and living environments, attracting a large number of high-end talent and innovation teams [70], which in turn attracts and gathers innovative talent and institutions, promotes the innovation level of the city itself, and creates a siphon effect on the talents of neighboring cities [58]. Due to the connections between various links in the industrial chain, the location choice of enterprises tends to be areas with spatial comparative advantages, thereby actively or passively influencing the innovation capacity and green total-factor productivity of neighboring cities. Therefore, the paper's third null hypothesis (H3a_0) and alternative hypothesis (H3a_1) are stated as follows:

Null hypothesis (H3a_0). *The implementation of smart-city policies has no significant spatial effects capable of influencing the innovation capacity and green total-factor productivity of neighboring cities through city network externalities.*

Alternative hypothesis (H3a_1). *The implementation of smart-city policies has spatial effects capable of influencing the innovation capacity and green total-factor productivity of neighboring cities through city network externalities.*

Furthermore, apart from city externalities, differences in city development inherently exist. Studies have pointed out that, compared to state-owned enterprises, enterprises in light-pollution industries, and enterprises in non-resource-based cities, the impact of green innovation on SCP is more significant for non-state-owned enterprises, enterprises in heavily polluting industries, and enterprises in non-resource-based cities [39]. Scholars have also found that the construction of smart cities helps enhance urban innovation performance. Still, this effect may vary among different cities, with each city exhibiting some unique outcomes. The impact of smart-city construction on CO_2 emission reductions

is more significant in cities with higher administrative levels, higher levels of neutral technological progress and green innovation, and more advanced industrial structures [43]. In this context, a neighboring city is defined as a neighboring city based on its geographical location. If it is neighboring, it is a neighboring city; if not, it is a nonneighboring city. Therefore, this paper's fourth null hypothesis (H3b_0) and alternative hypothesis (H3b_1) are stated as follows:

Null hypothesis (H3b_0). *The implementation of smart-city policies causes no significant regional or city differences in influencing urban innovation capacity and green total-factor productivity.*

Alternative hypothesis (H3b_1). *The implementation of smart-city policies causes regional and city differences in influencing urban innovation capacity and green total-factor productivity.*

3.2. Selection of Methods and Models

The difference-in-differences (DID) model [14], as one of the primary means of studying the net effects of policies, becomes particularly important when examining the differences in innovation capacity and sustainable development levels between smart pilot cities and non-pilot cities. Considering that the selection and establishment times of smart pilot cities exhibit significant variations across cities nationwide, this variation is further amplified under the influence of policies, thereby increasing the differences in innovation capacity and sustainable development levels between the two types of cities. Therefore, this study treats the designation of smart pilot cities as an approximate natural experiment for in-depth exploration. However, the non-random selection of smart pilot cities may lead to selection bias in the experiment, which could potentially cause endogeneity issues in the sample. The classic econometric model of difference-in-differences (DID) is employed to address the endogeneity issues caused due to selection bias. Since the approvals of smart pilot cities are not concentrated in a single year, this study uses a multi-period DID model for analysis. For multi-period panel data, the model is defined as follows:

$$y_{it} = \beta_0 + \beta_1 DID_{it} + X_{it}\gamma + \lambda_i + v_t + \epsilon_{it} \quad (1)$$

In which y_{it} represents the i-th city in the j year of the explained variable, which is mainly UCI, $GTFP$, GEC, or GTC in this paper, respectively; DID_{it} represents the i-th city in the j year of the dummy variable for the smart pilot-city policy; β_1 denotes the policy effect or treatment effect; λ_i stands for individual fixed effects; X_{it} are the control variables; v_t represents time fixed effects; ϵ_{it} is the disturbance term.

In the classical difference-in-differences (DID) method, there is a critical assumption called the "Stable Unit Treatment Value Assumption" (SUTVA). This assumption posits that the potential outcome of each unit should be unique after receiving treatment and not influenced by the treatment of other units [71]. However, in the real spatial dimension, it is challenging to consider each city as a completely independent spatial unit. Particularly when considering the implementation of smart pilot policies, the spatial correlation of these policies and potential spillover effects may render the SUTVA non-applicable in practice [72]. Furthermore, the transmission of spatial effects may be in not only the spatial lag terms of the dependent variable but also the variations in the error terms caused by random shocks [73,74]. Therefore, this study constructed a Difference-in-Differences Spatial Durbin Model (DID-SDM):

$$y_{it} = \beta_0 + \beta_1 DID_{it} + \delta_1 W_n * DID_{it} + \rho W_n * y_{it} + X_{it}\gamma + \lambda_i + v_t + (I - \lambda W_n)\epsilon_{it} \quad (2)$$

Among them, W_n is an $n * n$ spatial weight matrix where each element in the matrix represents the connection between region i and region j. Due to the existence of spillover effects, smart pilot cities may have an impact on neighboring cities, and this average spillover effect is measured by δ_1. Therefore, in the case of incorporating spatial spillover effects, the policy effect for the experimental group, and the policy effect for the control

group, is $\delta_1 W_n$. The ρ represents the impact of the explained variable of adjacent regions in the defined weight matrix. The other variables are the same as in Formula (1).

The classic difference-in-differences (DID) method requires that the experimental group and the control group satisfy the parallel trend assumption before the policy implementation, showing no significant difference before the policy and significant difference after the policy implementation. That is, the regression coefficients are not significant before the policy implementation, and they become significant after the policy implementation, thereby verifying the difference in the changing trend of the experimental group and the control group before and after the policy implementation. This study adopts the event study methodology of Beck et al. (2010) [75] to construct the parallel trend test model, as shown in the following formula:

$$y_{it} = \alpha_1 + \beta_{-5} time_{it}^{-5} + \beta_{-4} time_{it}^{-4} + \cdots + \beta_5 time_{it}^{5} + X_{it}\gamma + \lambda_i + v_t + \epsilon_{it} \quad (3)$$

Here, in the variable $time_{it}^{n}$, where the superscript $n > 0$ indicates n years after the baseline year of policy implementation, and $n < 0$ indicates n years before the baseline year of policy implementation. ϵ_{it} is the random disturbance term. The other variables are the same as in Formula (1).

4. Materials and Study Design

4.1. Materials' Data and Variables Description

According to the data published by the Ministry of Civil Affairs of the People's Republic of China, there are a total of 293 prefecture-level cities in China (http://xzqh.mca.gov.cn/map (accessed on 28 June 2024)). However, between 2007 and 2020, with the development of cities, some statistical indicators at the urban scale might be missing. For the convenience of this study, data from 15 prefecture-level cities with substantial missing indicators were removed. (The deleted prefecture-level cities include Tangshan City, Puer City, Tongren City, Bijie City, Danzhou City, Sansha City, Turpan City, Hami City, Haidong City, Lhasa City, Xigaze City, Qamdo City, Nyingchi City, Shannan City, and Naqu City). Ultimately, this study constructed a balanced panel dataset of 278 prefecture-level cities from 2007 to 2020, totaling 3892 samples.

The data required to construct the urban innovation capability index in this study mainly came from two parts of microdata: the number of patents published by the China National Intellectual Property Administration and the registered capital of enterprises from the China State Administration for Industry and Commerce [13]. Meanwhile, the data required to construct the green total-factor productivity index and the control-variable data were primarily sourced from the "China City Statistical Yearbook" (https://www.stats.gov.cn/zs/tjwh/tjkw/tjzl/202302/t20230215_1907995.html (accessed on 28 June 2024)), Wind database (https://www.wind.com.cn/mobile/EDB/zh.html (accessed on 28 June 2024)), EPS database (https://www.epsnet.com.cn/index.html (accessed on 28 June 2024)), statistical yearbooks officially released by provinces, socio-economic statistical bulletins of the provinces, and various databases or web pages of municipal government websites.

For part of the missing data in 2017–2018, interpolation was used to complete the data, ensuring its integrity and accuracy. As a result, a balanced panel dataset was ultimately obtained, including 132 cities that were selected as smart-city pilots and 146 cities that were not selected as smart-city pilots.

4.2. Explained Variables

The dependent variables are the urban innovation capability index (UCI) and sustainable development (green total-factor productivity, $GTFP$). The urban innovation capability index (UCI) was calculated based on the total-factor productivity, R&D investment, and the number of patents, following the calculation method in the "China Urban and Industrial Innovation Report 2017" [13]. Green total-factor productivity ($GTFP$) was calculated using indicators such as the number of employees in the city district, the built-up area of the city district, the capital stock at the 2006 base period, and the GDP deflator for the 2006

base period. The $GTFP$ was measured from the perspective of input–output to characterize the level of sustainable development in a city. The specific measurement process was as follows: (1) The EBM model proposed by Tone et al. (2010) [76] was used to calculate the environmental efficiency of urban development, which includes undesirable environmental outputs. (2) The GML index was used to decompose the results of the EBM model, yielding the dynamic changes in green total-factor productivity and its decomposition items. The decomposition items mainly include two parts: the green efficiency change (GEC) index and the green technical change (GTC) index. The green efficiency change index (GEC) reflects the impact on resource allocation efficiency due to changes in environmental factors resulting from policy and institutional changes. The green technical change index (GTC) refers to the outward shift of the production frontier, driven by technological development. Based on the characteristics of the data, an indicator system for measuring green total-factor productivity was constructed, with output indicators selected as shown in Table 1.

Table 1. Green total-factor productivity measurement indicator system.

Target Level	Normative Layer		Indicator Layer	Unit
Green total-factor productivity ($GTFP$)	Input indicators	Labor input	Number of persons employed in municipal districts	10,000 people
		Land input	Built-up area of municipal districts	Hm^2
		Capital investment	2006 base period capital stock	10,000 yuan (RMB)
	Output indicators	Expected outputs	GDP 2006 base period deflator	10,000 yuan (RMB)
		Indicators of undesired outputs	Sulfur dioxide SO_2	t
			Industrial wastewater	wt
			Soot	t

4.3. Core Explanatory Variable

This paper primarily investigates the impact of China's pilot smart-city policies on urban innovation and sustainable development. Therefore, the core explanatory variable of this study is whether a city is included in the pilot smart-city list, denoted as the variable DID. The first batch of pilot smart cities in China was established in 2013, so the period from 2007 to 2012 is considered the pre-experiment window for the pilot smart-city policy, and the period from 2013 to 2020 is the experimental period:

(1) The first batch of smart-city pilot websites in 2012: https://www.mohurd.gov.cn/gongkai/zhengce/zhengcefilelib/201212/20121204_212182.html (accessed on 28 June 2024);
(2) The second batch of smart-city pilot websites in 2013: https://www.mohurd.gov.cn/gongkai/zhengce/zhengcefilelib/201302/20130205_212789.html (accessed on 28 June 2024);
(3) The third batch of smart-city pilot websites in 2013: https://www.mohurd.gov.cn/gongkai/zhengce/zhengcefilelib/201308/20130805_214634.html (accessed on 28 June 2024);
(4) The fourth batch of smart-city pilot websites in 2015: https://www.mohurd.gov.cn/gongkai/zhengce/zhengcefilelib/201504/20150410_220653.html (accessed on 28 June 2024).

The sample of cities approved as pilot smart cities constituted the experimental group, while the other cities formed the control group. During the sample period, prefecture-level cities approved as pilot smart cities were assigned a value of 1 from the year they were approved, i.e., $DID = 1$, and 0 before approval, i.e., $DID = 0$; other non-pilot smart cities were consistently assigned a value of 0, i.e., $DID = 0$. Due to incomplete data for some cities, these were excluded; moreover, some control-group cities were also excluded due to the large number.Therefore, the final study includes 132 pilot smart cities as the experimental-group sample and 146 cities as the control-group sample.

4.4. Control Variables

Based on the existing research literature, the study controlled for urban innovation development and sustainable development by considering three main aspects: macroeconomic development, technological government support, and basic geographical characteristics. Macroeconomic development mainly includes urban per-capita GDP ($pgdp$) and urban employment conditions ($eoye$). Technological government support mainly includes regional government educational expenditure ($edexp$) and regional government scientific and technological expenditure ($sciexp$). The basic geographical characteristics of a city mainly include the urban built-up area (bua). In addition, to facilitate the robustness of the analytical conclusions, urban per-capita GDP ($pgdp$), urban employment conditions ($eoye$), and regional government educational expenditure ($edexp$) were substituted with regional gross domestic product (pdg), the proportion of the secondary industry in the regional gross domestic product (sis), and local general public budget expenditure ($fina$), respectively, in the robustness test. To unify the scales of variables and reduce multi-collinearity, the formula $ln(x+1)$ transformation was applied to all the control variables.

The descriptive statistics of each variable were carried out to intuitively understand the overall situation of each variable. Table 2 shows the definition and descriptive statistics of variables.

According to the descriptive statistical results of variable data, it can be found that the mean value, standard deviation, minimum value, and maximum value of the UCI during the years 2007 to 2020 were 12.84, 51.11, 0.0060, and 1309, respectively. The values of variable UCI differed significantly. Meanwhile, the mean value, standard deviation, minimum value, and maximum value of $GTFP$ were 1.002, 0.0101, 0.904, and 1.127; the value of $GTFP$ did not differ and was the same for GEC and GTC. From the perspective of the implementation of smart-city policies, the average UCI of cities that have implemented smart-city policies was 32.68, which is significantly higher than the average UCI of 5.355 for cities that have not implemented such policies. From a numerical perspective, there was not much difference between the cities' $GTFP$, GEC, and GTC before and after the implementation of the guidelines. To further discuss the spatial impact of the policies on UCI, $GTFP$, GEC, and GTC before and after their implementation, the subsequent part will proceed with regression analysis.

4.5. Weight Matrix

The spatial weight matrix W constructed in this study consists of an anti-geographic distance matrix, spatial adjacency matrix, and economic-distance matrix. The inverse geographical-distance weight matrix was obtained via the reciprocal of the Euclidean distance between city i and city j. The spatial-adjacency-weight matrix is mainly based on whether city i and city j are adjacent or not. The economic distance matrix measures the GDP gap between city i and city j. Because the geographic distance and the economic distance tend to produce problems of distance definition not unique and endogenous, this paper mainly selected the spatial adjacency matrix (W_n) as the spatial-weight matrix for analysis.

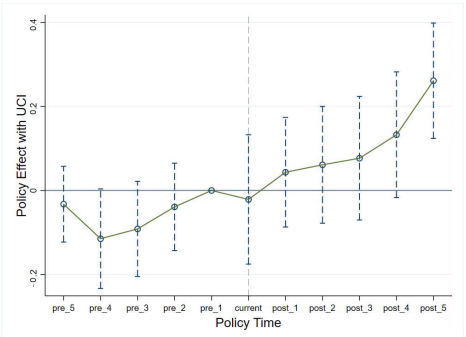
Table 2. Variables' selection, definition, and descriptive statistics.

	Sample Type	All Samples					$DID = 1$					$DID = 0$				
Variable	**Definition**	**n**	**Mean**	**Std.Dev**	**Min**	**Max**	**n**	**Mean**	**Std.Dev**	**Min**	**Max**	**n**	**Mean**	**Std.Dev**	**Min**	**Max**
UCI	The natural logarithm of innovation index	3892	12.84	51.11	0.00604	1309	1066	32.68	88.66	0.0962	1309	2826	5.355	20.73	0.00604	447.1
GTEP	The green total-factor productivity	3892	1.002	0.0101	0.904	1.127	1066	1.001	0.0144	0.904	1.127	2826	1.002	0.00791	0.911	1.084
GEC	The green technology efficiency index	3892	1.01	0.0315	0.882	1.143	1066	1.007	0.0414	0.882	1.143	2826	1.011	0.0268	0.892	1.131
GTC	The green technology progress index	3892	0.993	0.0311	0.867	1.139	1066	0.996	0.0421	0.867	1.139	2826	0.992	0.0257	0.88	1.106
DID	The policy dummy variable	3892	0.274	0.446	0	1										
edexp	The natural logarithm of education expenditure	3892	12.8	0.865	9.241	15.96	1066	13.35	0.768	10.91	15.96	2826	12.59	0.806	9.241	14.93
bua	The natural logarithm of the urban construction land area	3892	4.454	0.796	2.079	8.123	1066	4.869	0.83	2.708	7.209	2826	4.298	0.723	2.079	8.123
pgdp	The natural logarithm of the per-capita regional gross domestic product	3892	10.5	0.681	4.595	13.06	1066	10.94	0.561	9.084	13.06	2826	10.33	0.645	4.595	12.12
eoye	The natural logarithm of the number of employees in the unit at the end of the year	3892	12.73	0.8	2.97	15.74	1066	13.07	0.891	11.04	15.74	2826	12.6	0.722	2.97	15
sciexp	The natural logarithm of scientific and technological expenditures	3892	10	1.429	4.466	15.53	1066	10.94	1.392	7.297	15.53	2826	9.651	1.275	4.466	13.82
fina	The natural logarithm of local general public budget expenditure	3892	14.54	0.858	11.2	17.64	1066	15.14	0.731	12.86	17.64	2826	14.32	0.792	11.2	16.84
gdp	The natural logarithm of regional GDP	3892	7.133	0.943	4.141	10.23	1066	7.704	0.935	5.341	10.23	2826	6.917	0.852	4.141	9.604
sis	The natural logarithm of the proportion of the secondary industry in the gross regional domestic product	3892	3.846	0.255	2.458	4.521	1066	3.83	0.231	2.542	4.357	2826	3.852	0.264	2.458	4.521

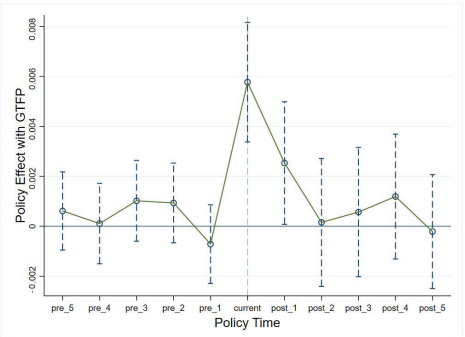
5. Parallel Trend Test, Empirical Results, and Discussion

5.1. Parallel Trend Test

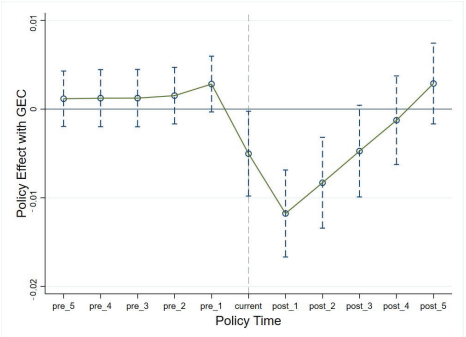
The classical differential method requires that the experimental group and the control group meet the parallel trend hypothesis before the implementation of the policy, that there is no significant difference before the implementation of the policy, and there is a significant difference after the implementation of the policy (that is, the regression coefficient before the implementation of the policy is not significant, but the regression coefficient is significant after the implementation of the policy) to verify whether the changing trend of the experimental group and the control group before and after the implementation of the policy is different. Therefore, for the relevant dependent variables of this paper, the urban innovation capability index (*UCI*) and green total-factor productivity (*GTFP*), a parallel trend test was conducted on the green efficiency change (*GEC*) and the green technical change (*GTC*). Figure 1a–d show the change in the regression coefficient before and after the implementation of smart pilot cities. The horizontal axis *post* is posted after the policy is implemented, *pre* is before the policy is implemented, and 0 is the base period.



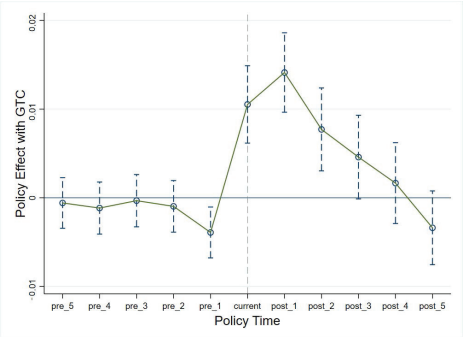
(**a**) The urban innovation capability index (*UCI*)



(**b**) The green total-factor productivity (*GTFP*)



(**c**) The green efficiency change (*GEC*)



(**d**) The green technical change (*GTC*)

Figure 1. Parallel trend test figures. Notes: The green line represents the line of the point estimates for the policy effect regression coefficient at each time point, while the blue dashed line represents the 95% confidence interval for the policy effect regression coefficient at each time point.

It can be observed from Figure 1a–d that pilot cities had no significant impact on any of the dependent variables (*UCI*, *GTFP*, *GEC*, and *GTC*) before the construction of smart-city policies. Still, after the implementation of smart-city policies, different dependent variables showed various effects. On the one hand, Figure 1a, Figure 1b, and Figure 1d, respectively, show that, after a smart-city pilot policy, the urban innovation capacity index, green total-factor productivity, and green technology changes showed an inevitable upward trend, which is reflected as a positive impact. In addition, the *GTFP* and *GTC* had a positive impact on policy implementation in the current period, with a time lag effect of one to three periods, while the *UCI* had no significant impact on policy implementation in the current period but a significant positive effect on the lag period of five periods. On the other hand,

Figure 1c shows a negative impact after the implementation of the smart-city pilot policy, which means the green efficiency change had a specific downward trend. It can be seen that all the samples studied in this paper met the parallel trend test.

5.2. Regression Results and Discussion

5.2.1. Basic Regressions

In this study, the traditional DID model was used as the benchmark regression model. The results in Table 3 show the regression results of the benchmark model, and the estimated results obtained using fixed effects and random effects are listed, respectively.

As shown in Table 3, through the Hausman test of fixed-effect and random-effect models, it was found that there were fixed effects in the regression models of the innovation index (*UCI*) and the green total-factor productivity (*GTFP*). However, in the regression results of the green efficiency change (*GEC*) and the green technical change (*GTC*), the fixed effect was not significant, so the random effect results could be selected for analysis. So, the policy-effect coefficient of the main explanatory variable innovation index (*UCI*) and the green technical change (*GTC*) were significantly positive, according to Model (2) and Models (7), indicating that the construction of smart pilot cities promotes the improvement of urban innovation capacity and the green technical change. However, the policy effect coefficient of the green total-factor productivity (*GTFP*) and the green efficiency change (*GEC*) did not significantly refer to Model (4) or Model (5), indicating that smart pilot-city construction has no significant effect on the sustainable development or the green efficiency change of a city. The above is only a preliminary conclusion, but the spatial effect of each variable was not taken into account. Therefore, the spatial effect of each variable needed to be added later for analysis.

Table 3. Classic DID baseline regressions.

Variables	*UCI*		*GTFP*	
Model	Model (1)	Model (2)	Model (3)	Model (4)
Effect	RE	FE	RE	FE
DID	0.4002 ***	0.4224 ***	0.0005	−0.0007
	(18.54)	(18.97)	(1.25)	(−1.28)
edexp	0.5107 ***	0.7000 ***	0.0008 **	0.0004
	(20.93)	(22.89)	(2.31)	(0.50)
bua	0.5444 ***	0.4627 ***	−0.0044 ***	−0.0053 ***
	(19.86)	(14.04)	(−11.96)	(−6.29)
pgdp	0.1028 ***	−0.0707 *	−0.0002	0.0003
	(3.63)	(−1.95)	(−0.61)	(0.34)
eoye	0.0052	−0.0203	0.0022 ***	0.0065 ***
	(0.22)	(−0.78)	(5.69)	(9.67)
sciexp	0.1435 ***	0.1173 ***	0	0.0003
	(9.92)	(7.75)	(−0.15)	(0.83)
Constant	−10.2962 ***	−9.9530 ***	0.9854 ***	0.9315 ***
	(−33.16)	(−29.33)	(195.91)	(107.16)
Fixedeffect	No	Yes	No	Yes
N	3892.00	3892.00	3892.00	3892.00
R^2	0.73	0.73	0.02	0.03
Hausmantest	chi2(6) = 101.80 (p = 0.0000)		chi2(6) = 69.85 (p = 0.0000)	

Table 3. *Cont.*

Variables	*GEC*		*GTC*	
Model	Model (5)	Model (6)	Model (7)	Model (8)
Effect	RE	FE	RE	FE
DID	−0.0017	−0.0046 **	0.0026 **	0.0043 **
	(−1.35)	(−2.44)	(2.05)	(2.29)
edexp	−0.0003	−0.0016	0.0017 *	0.002
	(−0.25)	(−0.62)	(1.66)	(0.79)
bua	−0.0043 ***	−0.0027	−0.0001	−0.0021
	(−3.80)	(−0.96)	(−0.05)	(−0.77)
pgdp	−0.0003	0.0012	0.0005	−0.0001
	(−0.31)	(0.39)	(0.50)	(−0.05)
eoye	0.0012	0.0062 ***	0.0006	−0.0001
	(0.97)	(2.77)	(0.53)	(−0.06)
sciexp	0.0004	0.0008	−0.0005	−0.0002
	(0.58)	(0.60)	(−0.66)	(−0.17)
Constant	1.0172 ***	0.9445 ***	0.9612 ***	0.9804 ***
	(64.86)	(32.66)	(61.99)	(34.19)
Fixedeffect	No	Yes	No	Yes
N	3892.00	3892.00	3892.00	3892.00
R^2	0.00	0.00	0.00	0.00
Hausmantest	chi2(6) = 10.71 (p = 0.0978)		chi2(6) = 2.28 (p = 0.8927)	

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, and the t value is in parentheses. When the p-value of the regression coefficient is less than 0.1, 0.05, or 0.01, it indicates that the null hypothesis can be rejected at the 10%, 5%, or 1% significance level, thus supporting the alternative hypothesis. However, if the p-value of the regression coefficient is greater than 0.1, it indicates that the null hypothesis cannot be rejected at any of the commonly used significance levels, thereby confirming that the null hypothesis is true. The same as below.

5.2.2. Spatial Effect Regressions

The traditional DID model may have the problem of missing variables, resulting in unrobust regression results. The above analysis shows that, due to the connection between cities, urban innovation efficiency and sustainable development levels do not change independently. A city's innovation efficiency and level of sustainable development may be influenced by the relevant policies or economic actions of other cities in the city network. Therefore, a DID-spatial Durbin model with spatial considerations is needed to more accurately estimate the impact of smart pilot-city policies on *UCI* and sustainable development. The DID-spatial Durbin model can control the influence of missing variables and consider the spatial spillover effect. By incorporating the spatial adjacency matrix, the model can capture the spatial interconnections between cities, thus obtaining more robust regression results. Therefore, this study used the DID-spatial Durbin model to analyze the impact of smart pilot-city policies on *UCI* and sustainable development. The premise of using the spatial metrology model is the existence of a spatial correlation among cities. In this study, Moran's I index test was conducted using the anti-geographic distance matrix to test whether there was a spatial effect on urban innovation efficiency. The results are shown in Table 4.

The Moran's I index of the *UCI*, *GTFP*, *GEC*, and *GTC* from 2007 to 2020 were almost significant and generally showed an increasing trend year by year, which basically could indicate that these variables all have a spatial correlation [75]. Therefore, the spatial econometric model could be used to study how the innovation capacity, green total-factor productivity, green efficiency change, and green technical change of cities may be affected by the relevant policies or economic behaviors of other cities in a city network.

Table 4. Moran's I test with *UCI*, *GTFP*, *GEC*, and *GTC*.

Variables	*UCI*	*GTFP*	*GEC*	*GTC*
2007	0.064 *	0.082 ***	0.125 ***	0.169 **
2008	0.097 **	0.167 ***	0.266 ***	0.245 ***
2009	0.142 ***	0.141 ***	0.237 ***	0.224 ***
2010	0.183 ***	0.159 ***	0.223 ***	0.220 ***
2011	0.221 ***	0.115 ***	0.113 ***	0.122 ***
2012	0.254 ***	0.079 **	0.144 ***	0.121 ***
2013	0.275 ***	0.088 **	0.391 ***	0.350 ***
2014	0.291 ***	−0.075 *	0.187 ***	0.376 ***
2015	0.312 ***	0.01	0.046	0.280 ***
2016	0.335 ***	0.058	0.015	0.066 *
2017	0.347 ***	0.118 ***	0.127 ***	0.183 ***
2018	0.363 ***	0.434 ***	0.281 ***	0.371 ***
2019	0.369 ***	0.218 ***	0.361 ***	0.378 ***
2020	0.380 ***	0.08 **	−0.019	0.169 ***

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, and the t value is in parentheses.

To further investigate the impact of policy actions on urban innovation efficiency and sustainable development, this study incorporated a spatial-adjacency matrix in the DID-spatial Durbin model, and the regression results are shown in Table 5. Overall, the regression results of the SDM model are consistent with those of the traditional DID regression, indicating that the SDM model is more robust.

Table 5. Results of SDM regression with spatial-adjacency matrix.

Variables	*UCI*		*GTEP*	
Model	Model (9)	Model (10)	Model (11)	Model (12)
DID	0.3609 ***	0.2796 ***	−0.0006	−2.0012 **
	(21.46)	(16.78)	(−2.47)	(−2.22)
$edexp$		0.2381 ***		0.0002
		(10.80)		(0.40)
bua		0.3312 ***		−2.0069 ***
		(13.84)		(−20.73)
$pgdp$		−2.1388 ***		0.0048 ***
		(−2.62)		(7.08)
$eoye$		0.0621 ***		0.0083 ***
		(3.35)		(14.46)
$sciexp$		0.0800 ***		−2.0007 **
		(7.35)		(−2.07)
$Constant$	0.1846 ***	−2.1885 ***	0.9980 ***	0.8368 ***
	(3.30)	(−24.73)	(741.45)	(93.34)
$W_n * DID$	0.1163 ***	−2.1008 ***	0.0009	−2.0029 ***
	(4.30)	(−2.56)	(1.42)	(−2.33)
ρ	0.7690 ***	0.6194 ***	0.0037 ***	0.0450 ***
	(84.97)	(51.10)	(2.70)	(13.04)
σ_e^2	0.0802 ***	0.0759 ***	0.0001 ***	0.0001 ***
	(40.99)	(40.65)	(42.48)	(41.85)
N	3892	3892	3892	3892
R^2	0.238	0.667	0.002	0.005
DID	−2.0048 ***	−2.0049 ***	0.0053 ***	0.0050 ***
	(−2.31)	(−2.28)	(5.35)	(4.84)

Table 5. *Cont.*

Variables	*GEC*		*GTC*	
Model	Model (13)	Model (14)	Model (15)	Model (16)
edexp		−2.0003		−2.0011
		(−2.18)		(−2.82)
bua		−2.0036 **		0.0006
		(−2.20)		(0.41)
pgdp		−2.0016		0.0048 ***
		(−2.91)		(2.99)
eoye		0.0045 ***		−2.0006
		(3.43)		(−2.48)
sciexp		0.0009		−2.0009
		(1.19)		(−2.40)
Constant	0.2826 ***	0.2529 ***	0.2362 ***	0.2136 ***
	(24.50)	(11.69)	(22.37)	(10.93)
$W_n * DID$	0.0048 ***	0.0056 ***	−2.0050 ***	−2.0069 ***
	(3.23)	(3.00)	(−2.81)	(−2.13)
ρ	0.7361 ***	0.7369 ***	0.7786 ***	0.7795 ***
	(76.40)	(76.69)	(92.72)	(92.94)
σ_e^2	0.0004 ***	0.0004 ***	0.0003 ***	0.0003 ***
	(40.21)	(40.20)	(40.34)	(40.33)
N	3892	3892	3892	3892
R^2	0.0000	0.0000	0.0000	0.0000

Notes: ** $p < 0.05$, *** $p < 0.01$, and the t value is in parentheses.

Based on the SDM regression results in Table 5, it was found that the impact of the smart pilot-city policy on *UCI* and *GTFP* was mainly reflected in two aspects. Firstly, the policy had a direct positive effect on innovation capacity, which was verified according to the significant positive coefficient of the policy dummy variable DID in Model (10), indicating that the smart pilot-city policy had a facilitating effect on the improvement of innovation capacity. This also verifies the reasonableness of H1_1. Policies have a direct negative impact on *GTFP*, which was verified according to the significant negative coefficient of the policy dummy variable DID. This indicates that smart pilot-city policies may inhibit the sustainable development of cities. This is inconsistent with the non-significant conclusion obtained using the panel regression Model (4) in Table 3, so the situation of H2_0 may occur. Secondly, there was a spillover effect of smart pilot-city policy on *UCI*, *GTFP*, *GEC*, and *GTC*. The coefficients of the spatial lag term ($W_n * DID$) of the policy dummy variables in models (10), (12), (14), and (16) were significantly negative or positive at the 5% level, and the coefficients that inhibit the spillover effect on *UCI*, *GTFP*, and *GTC* while having a positive spillover effect on *GEC* were all significantly negative or positive. This indicates that the construction of smart cities can effectively promote the *GEC* of neighboring cities, the policy of smart pilot cities has an "inhibitory" effect on the *UCI*, *GTFP*, and *GTC* of neighboring cities, and the policy of smart pilot cities can promote the sustainable development of neighboring cities. This verifies H3a_1.

The reason for this is that the construction of smart cities is often concentrated in one city, leading to a high concentration of resources and technologies. This concentration may put neighboring cities at a disadvantage in terms of access to resources and technological support, thereby limiting their innovative activities. For example, smart cities may attract large amounts of talent, capital, and technology, making it difficult for neighboring cities to access these resources in order to drive innovation. The development of smart cities may trigger a "catch-up effect" in neighboring cities, whereby neighboring cities have to invest more in similar developments to remain competitive. However, such catching up may not be based on their own development needs and realities but, rather, on blind imitation, leading to a waste of resources and inefficiency [77]. At the same time, neighboring

cities may face technical, financial, and managerial challenges in the process of catching up, further inhibiting their *UCI*, *GTFP*, and *GTC*. The construction of smart cities may intensify the competition between cities, leading to less cooperation between neighboring cities. This competitive relationship may result in a lack of cooperation mechanisms such as information sharing and technology exchange between cities, thus limiting the scope and depth of innovation activities. This makes the negative spatial spillover effect of the smart pilot-city policy on *UCI*, *GTFP*, and *GTC* more prominent. This also validates H3a_1 of the study.

5.3. Robustness Test

5.3.1. Variable Substitution

To check whether the SDM regression model is robust to various potential interfering factors, this study tried different combinations of control variables to observe whether the coefficients and significance levels of the core explanatory variables changed significantly, as shown in Table 6.

Table 6. Results of variable substitution regression.

Variables	*UCI*		*GTEP*	
Model	Model (9)	Model (10)	Model (11)	Model (12)
DID	0.3609 ***	0.2725 ***	−2.0006	−2.0002
	(21.46)	(16.68)	(−2.47)	(−2.46)
$fina$		−2.1921 ***		0.0032 ***
		(−2.39)		(4.56)
bua		0.3109 ***		−2.0019 ***
		(13.57)		(−2.30)
gdp		0.4727 ***		−2.0027 ***
		(13.30)		(−2.22)
sis		−2.7221 ***		0.0103 ***
		(−28.01)		(9.23)
$sciexp$		0.0953 ***		0.0010 ***
		(8.50)		(2.95)
$Constant$	0.1846 ***	0.4368	0.9980 ***	0.9031 ***
	(3.30)	(1.50)	(741.45)	(100.95)
$W_n * DID$	0.1163 ***	−2.1385 ***	0.0009	−2.0006
	(4.30)	(−2.96)	(1.42)	(−2.74)
ρ	0.7690 ***	0.5489 ***	0.0037 ***	0.0305 ***
	(84.97)	(42.12)	(2.70)	(12.39)
σ_e^2	0.0802 ***	0.0740 ***	0.0001 ***	0.0001 ***
	−20.99	−20.92	−22.48	−22.13
N	3892	3892	3892	3892
R^2	0.238	0.727	0.002	0.009
Variables	*GEC*		*GTC*	
Model	Model (13)	Model (14)	Model (15)	Model (16)
DID	−2.0048 ***	−2.0040 ***	0.0053 ***	0.0047 ***
	(−2.31)	(−2.53)	(5.35)	(4.55)
$fina$		0.0039 **		−2.0021
		(2.01)		(−2.18)
bua		−2.0019		0.0004
		(−2.16)		(0.30)
gdp		−2.0074 ***		0.0041 *
		(−2.80)		(1.72)

Table 6. *Cont.*

sis		0.0109 ***		−2.0095 ***
		(4.02)		(−2.90)
sciexp		0.0012		−2.0003
		(1.53)		(−2.45)
Constant	0.2826 ***	0.2350 ***	0.2362 ***	0.2776 ***
	(24.50)	(10.28)	(22.37)	(12.90)
$W_n * DID$	0.0048 ***	0.0077 ***	−2.0050 ***	−2.0077 ***
	(3.23)	(4.06)	(−2.81)	(−2.50)
ρ	0.7361 ***	0.7333 ***	0.7786 ***	0.7764 ***
	(76.40)	(75.67)	(92.72)	(91.94)
σ_e^2	0.0004 ***	0.0004 ***	0.0003 ***	0.0003 ***
	(40.21)	(40.22)	(40.34)	(40.33)
N	3892	3892	3892	3892
R^2	0.0000	0.0000	0.0000	0.0000

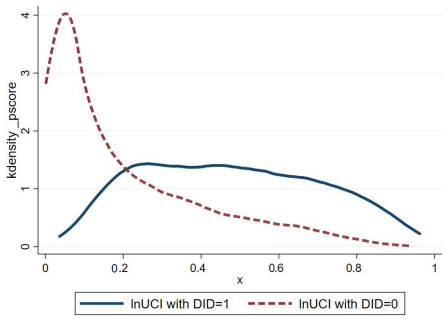
Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, and the t value is in parentheses.

The innovation capacity of smart pilot cities is influenced not only by the per-capita GDP of education expenditure but also by the financial expenditure, the proportion of the GDP of the whole city, and the GDP of secondary industries. Therefore, this study further analyzed the impact of smart pilot cities on *UCI*, *GTFP*, *GEC*, and *GTC* in the case of urban network externalities by rebuilding the spatial weight matrix based on the adjacency matrix of variables. In the stable bond regression results for *GTFP*, it was found that the impact of smart-city construction on urban *GTFP* is still not significant, which is consistent with the conclusion of Model (4) in Table 3 but inconsistent with the conclusion of Model (12) in Table 5, indicating that the impact of smart-city construction on the sustainable development of a city is not stable. In Table 5 of spatial regression results and Table 6 of steady-key regression results, the regression results of other variables are consistent except for differences in the impact of smart-city policies on *GTFP*. At the same time, it was found that the influence of smart-city policy construction on *GEC* and *GTC* were in opposite directions, which was precisely because of the effects of these two variables; the impact of a smart city on urban green total-factor productivity may not be significant. At the same time, the theoretical H1_1, H2_0, and H3a_1 proposed above were verified again.

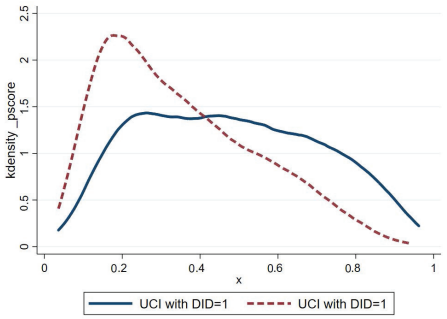
5.3.2. Propensity Score-Matching Method and Difference-in-Difference (PSM-DID)

The above analysis results show the following: (1) The construction of smart cities can enhance the innovation efficiency of a city. (2) The construction of smart cities has a spatial effect and can influence the *UCI* and *GTFP* of surrounding cities through urban network externalities. (3) The construction of smart cities may inhibit or have no significant effect on the *GTFP* of a city. However, the selection of which cities are to implement smart-city pilots is not random, and there may be a high level of *UCI* or a low level of *GTFP* in these cities, so the data may be biased. To further discuss the stability of the above conclusions, this section presents a regression test that was conducted on the impact of smart-city construction on *UCI* and *GTFP* based on PSM-DID. The propensity-matching score was mainly selected from one to four matching and kernel-matching methods, and the distribution diagram of the matched data was tested. The variable *UCI* was taken as an example to show its progress, as depicted in Figure 2.

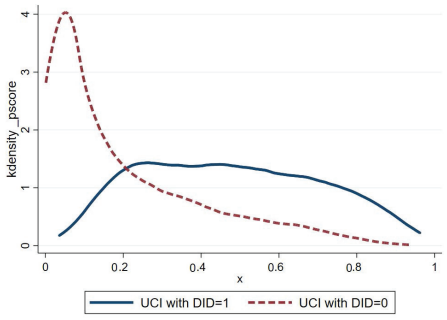
It can be found from the results of Figure 2, by comparing the probability density distribution maps before and after data matching, that the matching effect of one-to-four was revealed to be better than that of kernel-density matching, and both of them had a certain effect on the similarity of the data distribution. Based on the above matching data of one-to-four matching and nuclear matching, and combined with Formula (3), the effect of smart-city policy construction was analyzed, and the results are shown in Table 7.



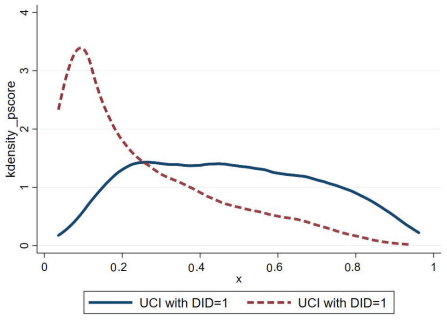
(**a**) Before one-to-four matching with *UCI*



(**b**) After one-to-four matching with *UCI*



(**c**) Before kernel matching with *UCI*



(**d**) After kernel matching with *UCI*

Figure 2. The density distribution of the *UCI* before and after matching.

Table 7. PSM-DID regression results.

Variable	*UCI*					
Matching	One-to-four matching			Kernel matching		
DID	1.2529 ***	0.6907 ***	0.4016 ***	30.0451 ***	23.2986 ***	22.1134 ***
	(29.46)	(14.98)	(15.45)	(19.93)	(13.00)	(12.97)
$W_n * DID$		0.2982 ***	−2.0075		3.3642 ***	0.1006
		(21.74)	(−2.84)		(6.84)	(0.20)
$W_n * UCI$			0.1772 ***			0.0932 ***
			(66.27)			(17.83)
Constant	1.2891 ***	0.9434 ***	0.1413 ***	4.1034 ***	0.9566	−2.0712
	(59.64)	(37.59)	(7.65)	(5.89)	(1.15)	(−2.09)
N	2241	2241	2241	3246	3246	3246
R^2	0.307	0.441	0.827	0.118	0.132	0.216
Variable	*GTFP*					
Matching	One-to-four matching			Kernel matching		
DID	−2.0007	0.0009	0.0008	−2.0006	0.0001	−2.0001
	(−2.75)	(0.73)	(0.72)	(−2.09)	(0.09)	(−2.11)
$W_n * DID$		−2.0009 **	−2.0006 *		−2.0003 *	−2.0001
		(−2.41)	(−2.87)		(−2.81)	(−2.59)
$W_n * GTFP$			0.1329 ***			0.1253 ***
			(18.60)			(22.14)
Constant	1.0022 ***	1.0032 ***	0.3534 ***	1.0021 ***	1.0024 ***	0.3907 ***
	(1979.69)	(1534.55)	(10.12)	(3799.15)	(3164.20)	(14.14)
N	2241	2241	2241	3246	3246	3246
R^2	0	0.003	0.153	0	0.002	0.143

Table 7. *Cont.*

Variable	*GEC*					
Matching	One-to-four matching			Kernel matching		
DID	−2.0089 ***	−2.0092 **	−2.0077 ***	−2.0066 ***	−2.0059 ***	−2.0061 ***
	(−2.76)	(−2.35)	(−2.64)	(−2.48)	(−2.60)	(−2.69)
$W_n * DID$		0.0001	0.0001		−2.0003	0.0007 *
		(0.11)	(0.17)		(−2.55)	(1.96)
$W_n * GEC$			0.1723 ***			0.1728 ***
			(68.57)			(77.14)
Constant	1.0131 ***	1.0130 ***	0.1638 ***	1.0125 ***	1.0129 ***	0.1614 ***
	(615.67)	(475.99)	(13.17)	(1157.79)	(963.80)	(14.60)
N	2241	2241	2241	3246	3246	3246
R^2	0.004	0.004	0.707	0.004	0.004	0.669
Variable	*GTC*					
Matching	One-to-four matching			Kernel matching		
DID	0.0094 ***	0.0108 ***	0.0089 ***	0.0072 ***	0.0067 ***	0.0065 ***
	(2.92)	(2.76)	(4.60)	(3.84)	(2.96)	(5.41)
$W_n * DID$		−2.0007	−2.0007		0.0003	−2.0008 **
		(−2.62)	(−2.23)		(0.45)	(−2.41)
$W_n * GTC$			0.1742 ***			0.1752 ***
			(77.09)			(87.19)
Constant	0.9899 ***	0.9908 ***	0.1474 ***	0.9903 ***	0.9901 ***	0.1451 ***
	(604.16)	(467.60)	(13.41)	(1139.90)	(948.35)	(14.94)
N	2241	2241	2241	3246	3246	3246
R^2	0.004	0.005	0.753	0.005	0.005	0.721

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, and the t value is in parentheses.

The following can be seen from the results in Table 7: (1) At the 1% level, smart-city construction had a significant positive impact on urban innovation (*UCI*), and urban innovation had a significant positive spatial spillover effect, while the policy space had a certain positive spatial spillover effect, but the effect was unstable, regardless of whether the spatial effect variables of policy and the urban innovation index were added. (2) Regardless of one-to-four matching or nuclear matching, in the regression results obtained, smart-city construction policies had no significant impact on the green total-factor productivity of cities. However, at the 10% level, smart-city construction had a significant negative spatial effect on the green total-factor productivity of neighboring cities. Moreover, urban green total-factor productivity (*GTFP*) had a significant positive spatial effect on the development of urban green total-factor productivity in neighboring cities. (3) At the 5% level, smart-city construction had a significant negative impact on urban green efficiency change (*GEC*). In addition, there were positive spatial spillover effects and positive policy spatial effects for the green efficiency change. (4) At the 5% level, smart-city construction had a significant positive impact on the green technical change (*GTC*), and there was a significant positive spatial spillover effect and negative policy spatial effect of the green technical change, but similarly. Thus, H1_1, H2_0, and H3a_1 were tested again.

6. Further Results and Discussion

6.1. Regional Difference Analysis

The impact of smart pilot-city policies on innovation and sustainable development is different in different regions. In this section of the study, the experimental data were divided into three regions, eastern, central, and western, according to China's regional divisions to analyze the differences in city regions and gain a more intuitive understanding of the impact of the smart pilot-city policy on different regions. The results are shown in Table 8.

Table 8. Results of SDM regression with spatial-adjacency matrix in three regions.

Variables	Eastern			
	UCI	*GTFP*	*GEC*	*GTC*
Model	Model (25)	Model (26)	Model (27)	Model (28)
DID	0.3442 ***	−2.0006	−2.0086 ***	0.0062 ***
	(11.20)	(−2.57)	(−2.64)	(2.90)
$W_n * DID$	−2.1123 **	0.0029 **	0.0016	−2.0037
	(−2.19)	(2.12)	(0.45)	(−2.16)
ρ	0.5539 ***	0.0003	0.6916 ***	0.7539 ***
	(25.26)	(0.15)	(39.84)	(51.52)
Controls	Yes	Yes	Yes	Yes
N	1344	1344	1344	1344
Cities num	96	96	96	96
R^2	0.745	0.078	0.000	0.001
Variables	Central			
	UCI	*GTFP*	*GEC*	*GTC*
Model	Model (29)	Model (30)	Model (31)	Model (32)
DID	0.1954 ***	0.0000	−2.0006	0.0005
	(7.72)	(0.04)	(−2.52)	(0.55)
$W_n * DID$	−2.0034	0.0013	0.0055 ***	−2.0036 **
	(−2.08)	(1.36)	(2.80)	(−2.15)
ρ	0.6265 ***	0.3866 ***	0.7777 ***	0.8073 ***
	(29.48)	(13.18)	(57.96)	(68.36)
Controls	Yes	Yes	Yes	Yes
N	1400	1400	1400	1400
Cities num	100	100	100	100
R^2	0.662	0.07	0.007	0.008
Variables	Western			
	UCI	*GTFP*	*GEC*	*GTC*
Model	Model (33)	Model (34)	Model (35)	Model (36)
DID	0.3828 ***	0.0011 **	−2.0062 ***	0.0075 ***
	(12.02)	(2.07)	(−2.91)	(3.89)
$W_n * DID$	−2.0775	0.0001	0.0073 **	−2.0094 ***
	(−2.55)	(0.14)	(2.33)	(−2.32)
ρ	0.3383 ***	0.0016	0.6326 ***	0.6698 ***
	(12.18)	(1.57)	(32.45)	(38.04)
Controls	Yes	Yes	Yes	Yes
N	1148	1148	1148	1148
Cities num	82	82	82	82
R^2	0.614	0.048	0.000	0.000

Notes: ** $p < 0.05$, *** $p < 0.01$, and the t value is in parentheses.

From the results of Table 8, it can be seen from models (25), (29), and (33) that the policy has a positive and significant impact and a positive spatial spillover effect on the innovation level of eastern, central, and western China; the average direct effect was 0.3442, 0.1954, and 0.3828, respectively. The average spatial spillover effects were 0.5539, 0.6265, and 0.3383, respectively. This indicates that the direct promotion effect of smart-city construction in the central region was the lowest, while the spatial spillover effect was the highest, and the direct promotion effect of smart-city construction in the western region was the highest, while the spatial spillover effect was the lowest. In addition, the spatial spillover effect of

smart-city policy effects was significantly negative only in the eastern region but not in the central and western areas.

At the same time, it can be seen from models (26), (30), and (34) that the impact of smart-city construction on urban green total-factor productivity (*GTFP*) was different in different regions. The western region had a significant positive effect, but the eastern and central regions had no significant impact. Only the central region had a significant positive spatial spillover effect on green total-factor productivity, and the eastern region had a considerable polity-positive spatial spillover effect on green total-factor productivity.

In addition, the results of the remaining regression results revealed that the eastern and western regions had a significant effect on the change in the green effect and green technology change, in which the former was significantly negative, the latter was significantly positive, and there was a significant policy spatial effect in the central and western regions. In other words, the implementation of policies in cities of the western and central areas would promote the green effect or inhibit the green technology of cities in the neighboring regions, respectively. On the whole, there were significant positive spatial spillover effects for the urban green effect and green technology in the eastern, central, and western regions. This tested H3b_1.

To further analyze the impact of policy effects on *UCI* and *GTFP* in different regions, the region was further divided into seven regions: northeast, north, central, south, east, northwest, and southwest, and the regression results are shown in Table 9.

Table 9. Results of SDM regression with spatial-adjacency matrix in seven regions.

Variables	Northeast			
	UCI	*GTFP*	*GEC*	*GTC*
Model	Model (37)	Model (38)	Model (39)	Model (40)
DID	0.2346 ***	−2.0015	−2.001	0.0006
	(4.87)	(−2.77)	(−2.26)	(0.21)
$W_n * DID$	0.014	−2.0106 ***	0.0005	−2.0049
	(0.18)	(−2.29)	(0.10)	(−2.02)
ρ	0.4765 ***	0.0149 ***	0.6522 ***	0.7610 ***
	(11.20)	(3.87)	(19.17)	(30.49)
Controls	Yes	Yes	Yes	Yes
N	504	504	504	504
Cities num	36	36	36	36
R^2	0.737	0.02	0.001	0.002
Variables	North			
	UCI	*GTFP*	*GEC*	*GTC*
Model	Model (41)	Model (42)	Model (43)	Model (44)
DID	0.4262 ***	−2.0039 ***	−2.0084 *	0.0075 *
	(12.64)	(−2.89)	(−2.92)	(1.86)
$W_n * DID$	−2.2565 ***	−2.0035 *	0.006	−2.0089
	(−2.38)	(−2.70)	(0.88)	(−2.41)
ρ	0.5133 ***	0.0136 **	0.6555 ***	0.7120 ***
	(13.41)	(2.52)	(20.67)	(25.73)
Controls	Yes	Yes	Yes	Yes
N	420	420	420	420
Cities num	30	30	30	30
R^2	0.519	0.015	0	0

Table 9. *Cont.*

Variables	Midland			
	UCI	*GTFP*	*GEC*	*GTC*
Model	Model (45)	Model (46)	Model (47)	Model (48)
DID	0.2181 ***	0.0001	−2.0018	0.0019
	(5.70)	(0.12)	(−2.95)	(1.10)
$W_n * DID$	0.1247 *	0.0032 **	0.0072 *	−2.0034
	(1.68)	(2.20)	(1.94)	(−2.01)
ρ	0.5661 ***	0.4289 ***	0.7788 ***	0.7906 ***
	(15.95)	(9.96)	(40.73)	(43.68)
Controls	Yes	Yes	Yes	Yes
N	588	588	588	588
Cities num	42	42	42	42
R^2	0.69	0.121	0.009	0.01
Variables	South			
	UCI	*GTFP*	*GEC*	*GTC*
Model	Model (49)	Model (50)	Model (51)	Model (52)
DID	0.3743 ***	0.0011	−2.0009	0.0018
	(6.73)	(0.60)	(−2.38)	(0.90)
$W_n * DID$	−2.1117	0.0064 **	0.0114 ***	−2.005
	(−2.15)	(2.12)	(2.95)	(−2.43)
ρ	0.6163 ***	0.0804	0.6994 ***	0.7462 ***
	(17.71)	(1.39)	(25.36)	(29.72)
Controls	Yes	Yes	Yes	Yes
N	490	490	490	490
Cities num	35	35	35	35
R^2	0.807	0.087	0.035	0.015
Variables	Eastland			
	UCI	*GTFP*	*GEC*	*GTC*
Model	Model (53)	Model (54)	Model (55)	Model (56)
DID	0.1778 ***	−2.0012	−2.0056 **	0.0051 ***
	(5.89)	(−2.04)	(−2.47)	(2.60)
$W_n * DID$	−2.2040 ***	0.0001	0.0048	−2.0047
	(−2.21)	(0.07)	(1.38)	(−2.55)
ρ	0.6952 ***	0.1164 ***	0.7740 ***	0.8202 ***
	(34.42)	(8.52)	(46.43)	(58.20)
Controls	Yes	Yes	Yes	Yes
N	1064	1064	1064	1064
Cities num	76	76	76	76
R^2	0.682	0.008	0	0

Table 9. *Cont.*

Variables	Northwest			
	UCI	*GTFP*	*GEC*	*GTC*
Model	Model (57)	Model (58)	Model (59)	Model (60)
DID	0.2975 ***	0.0015 *	−2.0037	0.0051 *
	(6.36)	(1.91)	(−2.15)	(1.83)
$W_n * DID$	−2.1569 *	0.0012	0.0088 *	−2.0116 **
	(−2.91)	(0.92)	(1.68)	(−2.52)
ρ	0.1984 ***	0.0011	0.6419 ***	0.7097 ***
	(3.52)	(0.79)	(17.64)	(22.67)
Controls	Yes	Yes	Yes	Yes
N	420	420	420	420
Cities num	30	30	30	30
R^2	0.653	0.094	0.001	0
Variables	Southwest			
	UCI	*GTFP*	*GEC*	*GTC*
Model	Model (61)	Model (62)	Model (63)	Model (64)
DID	0.2581 ***	0.0013	−2.002	0.0031
	(4.14)	(1.18)	(−2.78)	(1.36)
$W_n * DID$	0.1692 *	0.0026	0.0067 *	−2.0046
	(1.76)	(1.63)	(1.74)	(−2.37)
ρ	0.2481 ***	0.2723 ***	0.6715 ***	0.6893 ***
	(5.56)	(5.71)	(24.65)	(26.67)
Controls	Yes	Yes	Yes	Yes
N	406	406	406	406
Cities num	29	29	29	29
R^2	0.722	0.052	0.011	0.004

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, and the t value is in parentheses.

According to models (37), (41), (45) (49), (53), (57), and (61), in Table 9, the regression results show that, among the regression results of *UCI*, there were significant positive direct policy effects and indirect spatial spillover effects in different regions, among which the direct policy effects were 0.2346, 0.4262, 0.2181, 0.3743, 0.1778, 0.2975, and 0.2581, respectively. The direct policy average effect in the midland region was the largest, followed by that in the south region. The east region had the smallest indirect spatial spillover effect, and the indirect spatial spillover effect was 0.4765, 0.5133, 0.5661, 0.6163, 0.6952, 0.198, and 0.2481, respectively. The east region had the most significant indirect spatial spillover effect, followed by the south region and northwest region. However, the spatial impacts of policies in the different areas were different, among which, the north, east, and northwest were significantly negative, the midland and southwest were significantly positive, and the northeast and south were not significant.

Similarly, from models (38), (42), (46) (50), (54), (58), and (62) in Table 9, it can be found that smart-city policies had no significant impact on urban green total-factor productivity in most regions. Only the northwest region had a significant positive effect at the 10% level. As for the regression results of the indirect spatial spillover effect and policy-spatial effect, it can be found that spatial effects were different in different regions. *GTFP* had positive spatial spillover effects in the northeast, north, midland, east, and southwest regions. There was a negative spillover effect of the policy space in the northeast and north regions and a positive spillover effect in the midland and south regions.

From the remaining regression results in Table 9, it can be seen that the implementation of smart-city policies had a different significance on urban *GEC* and *GTC*; some regions were significant, while others were not, but the symbols of influence on *GEC* and *GTC*

were the opposite. At the same time, both *GEC* and *GTC* had significant positive spatial spillover effects on neighboring cities, while there were few areas with significant spillover effects of policies. This also verified H3b_1.

6.2. City Difference Analysis

The previous analysis from the perspective of different geographical regions in China indicates that there are regional differences in the impact of smart-city construction on urban innovation capability (*UCI*), green total-factor productivity (*GTEP*), green efficiency (*GEC*), and green technology progress (*GTC*). In addition to the regional disparities, the development status of different cities could also have affected the outcomes of this study. Therefore, this paper further categorized the development levels of cities based on the annual regional GDP quartiles: the lower quartile (*Q*25), the median (*Q*50), and the upper quartile (*Q*75). The categorization rules were as follows: if a city's annual GDP was greater than or equal to the regional annual GDP's upper quartile (*Q*75), the city was defined as a first-tier city for that year. If a city's annual GDP was greater than or equal to the regional annual GDP median (*Q*50) but less than the regional annual GDP's upper quartile (*Q*75), the city was defined as a second-tier city for that year. If a city's annual GDP was greater than or equal to the regional annual GDP's lower quartile (*Q*25) but less than the regional annual GDP median (*Q*50), the city was defined as a third-tier city for that year. If a city's annual GDP was less than the regional annual GDP's lower quartile (*Q*25), the city was defined as a fourth-tier city for that year. The regression results based on samples divided into different urban regions are shown in Table 10.

Table 10. Results of SDM regression with spatial-adjacency matrix in four different city regions.

City	First-tier cities			
Variables	*UCI*	*GTFP*	*GEC*	*GTC*
Model	Model (65)	Model (66)	Model (67)	Model (68)
DID	0.2982 ***	−2.002	−2.0110 ***	0.0096 ***
	(8.92)	(−2.19)	(−2.66)	(3.55)
$W_n * DID$	0.0372 ***	−2.0011 **	−2.0002	−2.0011
	(3.44)	(−2.03)	(−2.25)	(−2.21)
ρ	0.0013 ***	0.1455 ***	0.1907 ***	0.1947 ***
	(17.23)	(11.37)	(39.49)	(46.46)
Controls	Yes	Yes	Yes	Yes
N	966	966	966	966
Cities num	69	69	69	69
R^2	0.929	0.239	0.65	0.723
City	Second-tier cities			
Variables	*UCI*	*GTFP*	*GEC*	*GTC*
Model	Model (69)	Model (70)	Model (71)	Model (72)
DID	0.3456 ***	−2.001	−2.0079 ***	0.0069 ***
	(9.39)	(−2.96)	(−2.21)	(3.09)
$W_n * DID$	0.0251 **	−2.0005 *	0.0005	−2.001
	(2.28)	(−2.66)	(0.71)	(−2.53)
ρ	0.0027 ***	0.1046 ***	0.1718 ***	0.1751 ***
	(19.56)	(12.18)	(48.15)	(53.27)
Controls	Yes	Yes	Yes	Yes
N	980	980	980	980
Cities num	70	70	70	70
R^2	0.848	0.209	0.731	0.769

Table 10. *Cont.*

Variables	Third-tier cities			
	UCI	*GTFP*	*GEC*	*GTC*
Model	Model (73)	Model (74)	Model (75)	Model (76)
DID	0.1838 ***	0.0008	0.0009	0.0006
	(4.85)	(0.85)	(0.37)	(0.28)
$W_n * DID$	0.0629 ***	−2.0005 *	0.001	−2.0016 ***
	(5.76)	(−2.71)	(1.55)	(−2.65)
ρ	0.0019 ***	0.0846 ***	0.1556 ***	0.1552 ***
	(13.03)	(9.11)	(40.91)	(45.63)
Controls	Yes	Yes	Yes	Yes
N	980	980	980	980
Cities num	70	70	70	70
R^2	0.761	0.163	0.666	0.711
City	Fourth-tier cities			
Variables	*UCI*	*GTFP*	*GEC*	*GTC*
Model	Model (77)	Model (78)	Model (79)	Model (80)
DID	0.1607 ***	−2.0005	−2.0007	0.0007
	(6.33)	(−2.82)	(−2.33)	(0.35)
$W_n * DID$	0.0431 ***	0.0002	0.0024 ***	−2.0022 ***
	(5.36)	(0.81)	(3.73)	(−2.66)
ρ	0.0032 ***	0.0697 ***	0.1593 ***	0.1582 ***
	(17.17)	(10.28)	(35.66)	(39.32)
Controls	Yes	Yes	Yes	Yes
N	966	966	966	966
Cities num	69	69	69	69
R^2	0.724	0.152	0.607	0.649

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$, and the t value is in parentheses.

According to the regression results of *UCI*, in models (65), (69), (73), and (77) in Table 10, it can be seen that smart-city policies had significant positive direct and indirect impacts on the innovation capability of different cities. The average direct impact effect was 0.2982, 0.3456, 0.1838, and 0.1607, respectively. First-tier and second-tier cities had a more significant impact, while third-tier and fourth-tier cities had a more minor impact. This shows that the construction of smart cities has a more noticeable effect on the innovation capacity of cities with sound economic development. At the same time, the spatial indirect impact of policies was 0.0372, 0.0251, 0.0629, and 0.0431, respectively. Among them, the influence of first-tier and second-tier cities was relatively small. In contrast, the impact of third-tier and fourth-tier cities was more significant, which was the opposite of the direct-effect size distribution. This indicates that the smart-city construction policy of a city with weak economic development has a more obvious promoting effect on the innovation capacity of neighboring cities.

According to the regression results of *GTFP*, in models (66), (70), (74), and (78) in Table 10, smart-city policies had no significant impact on urban green total-factor productivity. Meanwhile, the spatial effect of urban green total-factor productivity was significantly positive. However, in first-tier, second-tier, and third-tier cities, smart-city policies had a significant negative spatial effect on urban green total-factor productivity at the 10% level.

According to the regression results of *GEC* and *GTC* in models (67), (68), (71), (72), (75), (76), (79), and (80) in Table 10, it can be seen that the impact of smart-city policies on urban green efficiency or urban technological progress was significantly negative or positive in first-tier and second-tier cities, respectively. However, it was not significant in third-tier and fourth-tier cities. That is, there were also regional differences in economic

development. At the same time, the spatial effect of urban green efficiency or urban technological progress was significantly positive. However, in fourth-tier cities, smart-city policies had significant positive or negative policy-space spillover effects on urban green efficiency or urban technological progress at the 1% level, respectively. In contrast, in third-tier cities, smart-city policies only had significant adverse policy-space spillover effects on urban technological progress.

According to the value of regression coefficient ρ in Table 10, at the 1% level, urban innovation capability (*UCI*), urban green productivity (*GTEP*), urban green efficiency (*GEC*), and green technology progress (*GTC*) all had significant positive spatial spillover effects in the four metropolitan regions. Most of them showed the following rule: the better the urban regional economic development, the greater the average spillover effect. This also tested H3b_1.

7. Conclusions, Policy Recommendations, and Further Research

7.1. Main Conclusions

Urban innovation capacity and urban green total-factor productivity are the keys to sustainable urban development. To study the impact of smart-city policies on urban innovation capacity and green total-factor productivity, this paper has analyzed the direct and spatial indirect effects of China's smart-city pilot policies on urban innovation and the total productivity of urban green factors based on the quasi-natural experiment of China's smart-city pilot policies and 278 prefecture-level cities' data from 2007 to 2020. The main conclusions are as follows.

Firstly, the construction of smart pilot cities can significantly improve urban innovation capacity, but it has a negative spillover effect on innovation capacity.

Secondly, the impact of smart-city policies on urban green total-factor productivity is unstable or even insignificant because smart-city policies can inhibit the change in urban green efficiency (*GEC*) but can also promote the progress of urban green technology (*GTC*).

Thirdly, due to the existence of urban network externalities, the construction of smart pilot cities will have a negative spillover effect on the green total-factor productivity (*GTFP*) and the green technical change (*GTC*) of neighboring cities like the innovation capacity (*UCI*), but it will also have a positive spillover effect on the green efficiency change (*GEC*).

Fourthly, there are significant regional and city differences in the impact of implementing smart-city policies on urban innovation capacity, green total-factor productivity, green efficiency change, and green technology progress.

7.2. Policy Recommendations

Based on the research conclusions, the implications for the construction of smart pilot cities in China are as follows.

- We should further increase investment in technological research and development in smart pilot cities, actively attract and cultivate technological talent, and pool research resources to promote rapid progress in green technology.
- It is necessary to ensure that the EASTERN, CENTRAL, AND WESTERN regions all have smart cities with innovative potential and provide balanced support to enhance regional innovation capabilities.
- To create a favorable urban sustainable innovation environment, we need to start from multiple aspects. This includes learning from the construction experience of smart cities, improving intellectual property protection and talent-training systems, and optimizing infrastructure for transportation, research, education, and manufacturing.

7.3. Limitations of This Study and Further Research

This paper mainly analyzed the impact of China's smart-city pilot policy implementation on urban innovation and urban green total-factor productivity from the perspective of spatial effects. Further, it discussed the implications of green total-factor productivity on green efficiency change and green technology progress. However, this study did not

consider the spatial effect of selected smart cities on neighboring smart cities that were not chosen or the impact of non-selected smart cities on selected smart cities. Additionally, this study did not account for the channels through which smart-city implementation policies impact urban innovation capacity and sustainable urban development. Furthermore, the potential asymmetry in spatial effects between cities was not thoroughly examined in this study. These represent the limitations of our research, and they are subjects worthy of further investigation. We will address these issues in greater depth in future research.

Author Contributions: Conceptualization, S.Y.; methodology, S.Y.; software, S.Y. and Y.S.; validation, S.Y. and Y.S.; formal analysis, S.Y. and Y.S.; investigation, S.Y.; resources, S.Y.; data curation, S.Y.; writing—original draft preparation, Y.S.; writing—review and editing, S.Y., Y.S. and Q.Y.; visualization, S.Y. and Y.S.; supervision, S.Y.; project administration, S.Y.; funding acquisition, S.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the University Humanities and Social Science Research Project of Guizhou Education Department "Research on Digital Economy Empowering Rural Revitalization in Guizhou" (2024RW146).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analysed in this study. These data can be found here: https://pan.baidu.com/s/1MtL3d2kIVneHleg8cdQtIw?pwd=jje6 (code: jje6), accessed on 28 June 2024.

Acknowledgments: The authors thank the editors and reviewers for their comments, which led to the improvement of this paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Angelidou, M. Smart city policies: A spatial approach. *Cities* **2014**, *41*, S3–S11. [CrossRef]
2. Los Angeles Community Analysis Bureau (LACAB). *The State of the City: The Cluster Analysis of Los Angeles*; Community Analysis Bureau: Los Angeles, CA, USA, 1974.
3. Bakici, T.; Almirall, E.; Wareham, J. A smart city initiative: The case of Barcelona. *J. Knowl. Econ.* **2013**, *4*, 135–148. [CrossRef]
4. Nam, T.; Pardo, T.A. Conceptualizing smart city with dimensions of technology, people, and institutions. In Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times, College Park, MD, USA, 12–15 June 2011; Volume 4, pp. 282–291. [CrossRef]
5. Alawadhi, S.; Aldama-Nalda, A.; Chourabi, H.; Gil-Garcia, J.R.; Leung, S.; Mellouli, S.; Nam, T.; Pardo, T.A.; Scholl, H.J.; Walker, S. Building Understanding of Smart City Initiatives. In Proceedings of the 11th International Conference on Electronic Government (EGOV), Kristiansand, Norway, 3–6 September 2012; pp. 40–53. Available online: https://link.springer.com/chapter/10.1007/978-3-642-33489-4_4 (accessed on 28 June 2024).
6. Zubizarreta, I.; Seravalli, A.; Arrizabalaga, S. Smart city concept: What it is and what it should be. *J. Urban Plan. Dev.* **2016**, *142*, 4015005. [CrossRef]
7. Vanolo, A. Smart mentality: The smart city as disciplinary strategy. *Urban Stud.* **2014**, *51*, 883–898. [CrossRef]
8. Yigitcanlar, T.; Kankanamge, N.; Vella, K. How Are Smart City Concepts and Technologies Perceived and Utilized? A Systematic Geo-Twitter Analysis of Smart Cities in Australia. *J. Urban Technol.* **2021**, *28*, 135–154. [CrossRef]
9. Rodríguez-Bolívar, P.M. *Transforming City Governments for Successful Smart Cities*; Springer International Publishing: Berlin/Heidelberg, Germany, 2015; pp. 9–12. [CrossRef]
10. Lai, C.S.; Jia, Y.; Dong, Z.; Wang, D.; Tao, Y.; Lai, Q.H.; Wong, R.T.; Zobaa, A.F.; Wu, R.; Lai, L.L. A review of technical standards for smart cities. *Clean Technol.* **2020**, *2*, 290–310. [CrossRef]
11. Wang, K.-L.; Pang, S.; Zhang, F.; Miao, Z.; Sun, H. The impact assessment of smart city policy on urban green total-factor productivity: Evidence from China. *Environ. Impact Assess. Rev.* **2022**, *94*, 106756. [CrossRef]
12. Stübinger, J.; Schneider, L. Understanding smart city—A data-driven literature review. *Sustainability* **2020**, *12*, 8460. 10.3390/su12208460. [CrossRef]
13. Holgersson, T.; Kekezi, O. Towards a multivariate innovation index. *Econ. Innov. New Technol.* **2017**, *27*, 254–272. [CrossRef]
14. Zhong, J.; Chun, W.; Deng, W.; Gao, H. Can Mergers and Acquisitions Promote Technological Innovation in the New Energy Industry? An Empirical Analysis Based on China's Lithium Battery Industry. *Sustainability* **2023**, *15*, 12136. [CrossRef]
15. Ren, Y.; Fang, C.; Li, G. Spatiotemporal characteristics and influential factors of eco-efficiency in Chinese prefecture-level cities: A spatial panel econometric analysis. *J. Clean. Prod.* **2020**, *260*, 120787–120787. [CrossRef]

16. Hollands, R.G. Will The Real Smart City Please stand Up? Intelligent, Progressive Or Entrepreneurial? In *The Routledge Companion to Smart Cities*; Routledge: Abingdon, UK, 2008; Volume 21. [CrossRef]
17. Washburn, D.; Sindhu, U.; Balaouras, S.; Dines, R.A.; Hayes, N.M.; Nelson, L.E. *Helping CIOs Understand "Smart City" Initiatives: Defining the Smart City, Its Drivers, and the Role of the CIO*; Forrester Research, Inc.: Cambridge, MA, USA, 2010. Available online: http://public.dhe.ibm.com/partnerworld/pub/smb/smarterplanet/forr_help_cios_und_smart_city_initiatives.pdf (accessed on 28 June 2024).
18. Rios, P. Creating the Smart City. 2008. Available online: http://dspace.udmercy.edu:8080/dspace/bitstream/10429/20/1/2008_rios_smart.pdf (accessed on 28 June 2024).
19. Marsa-Maestre, I.; Lopez-Carmona, M.A.; Velasco, J.R.; Navarro, A. Mobile agents for service personalization in smart environments. *J. Netw.* **2008**, *3*, 30–41. [CrossRef]
20. Anthopoulos, L.; Tsoukalas, I.A. The Implementation Model of a Digital City. The Case Study of the Digital City of Trikala, Greece: E-Trikala. *J. e-Gov.* **2005**, *2*, 91–109. Available online: https://www.academia.edu/download/30696690/icfinal.pdf (accessed on 28 June 2024). [CrossRef]
21. Komninos, N. Intelligent cities: Towards interactive and global innovation environments. *Int. J. Innov. Reg. Dev.* **2009**, *1*, 337–355. [CrossRef]
22. Sairamesh, J.; Lee, A.; Anania, L. Session details: Information cities. *Commun. Acm.* **2004**, *47*, 28–31.10. [CrossRef]
23. Hall, P. Creative cities and economic development. *Urban Stud.* **2000**, *37*, 633–649. [CrossRef]
24. Coe, A.; Paquet, G.; Roy, J. E-Governance and Smart Communities. *Soc. Sci. Comput. Rev.* **2001**, *19*, 80–93. [CrossRef]
25. Dameri, R.P. Searching for Smart City definition: A comprehensive proposal. *Int. J. Comput. Technol.* **2013**, *11*, 2544–2551. [CrossRef]
26. Kim, T.; Ramos, C.; Mohammed, S. Smart city and IoT. *Future Gener. Comput. Syst.* **2017**, *76*, 159–162. [CrossRef]
27. Kirimtat, A.; Krejcar, O.; Kertesz, A.; Tasgetiren, M.F. Future trends and current state of smart city concepts: A survey. *IEEE Access* **2020**, *8*, 86448–86467. [CrossRef]
28. Neirotti, P.; De Marco, A.; Cagliano, A.C.; Mangano, G.; Scorrano, F. Current trends in Smart City initiatives: Some stylised facts. *Cities* **2014**, *38*, 25–36. [CrossRef]
29. Kim, J. Smart city trends: A focus on 5 countries and 15 companies. *Cities* **2022**, *123*, 103551. [CrossRef]
30. Rejeb, A.; Rejeb, K.; Simske, S.; Treiblmaier, H.; Zailani, S. The big picture on the internet of things and the smart city: A review of what we know and what we need to know. *Internet Thing* **2022**, *19*. 100565. [CrossRef]
31. Yaqoob, I.; Salah, K.; Jayaraman, R.; Omar, M. Metaverse applications in smart cities: Enabling technologies, opportunities, challenges, and future directions. *Internet Things* **2023**, *23*, 100884. [CrossRef]
32. Alizadeh, H.; Sharifi, A. Toward a societal smart city: Clarifying the social justice dimension of smart cities. *Sustain. Cities Soc.* **2023**, *95*, 104612. [CrossRef]
33. Kruhlov, V.; Dvorak, J.; Moroz, V.; Tereshchenko, D. Revitalizing Ukrainian Cities: The Role of Public-Private Partnerships in Smart Urban Development. *Cent. Eur. Public Adm. Rev.* **2024**, *22*, 85–107. [CrossRef]
34. Chen, M.A. Smart city and cyber-security; technologies used, leading challenges and future recommendations. *Energy Rep.* **2021**, *7*, 7999–8012. [CrossRef]
35. Mora, L.; Gerli, P.; Ardito, L.; Petruzzelli, A.M. Smart city governance from an innovation management perspective: Theoretical framing, review of current practices, and future research agenda. *Technovation* **2023**, *123*. 102717. [CrossRef]
36. Zhao, F.; Fashola, O.I.; Olarewaju, T.I.; Onwumere, I. Smart city research: A holistic and state-of-the-art literature review. *Cities* **2021**, *119*, 103406. [CrossRef]
37. Sharif, A.R.; Pokharel, S. Smart city dimensions and associated risks: Review of literature. *Sustain. Cities Soc.* **2022**, *77*, 103542. [CrossRef]
38. Bibri, S.E.; Krogstie, J. The emerging data–driven Smart City and its innovative applied solutions for sustainability: The cases of London and Barcelona. *Energy Inform.* **2020**, *3*, 26–471. [CrossRef]
39. Yang, S.B.; Jahanger, A.; Usman, M. Examining the influence of green innovations in industrial enterprises on China's smart city development. *Technol. Forecast. Soc. Chang.* **2024**, *199*, 123031. [CrossRef]
40. Gohar, A.; Nencioni, G. The role of 5G technologies in a smart city: The case for intelligent transportation system. *Sustainability* **2021**, *13*, 5188. [CrossRef]
41. Alahi, M.E.E.; Sukkuea, A.; Tina, F.W.; Nag, A.; Kurdthongmee, W.; Suwannarat, K.; Mukhopadhyay, S.C. Integration of IoT-enabled technologies and artificial intelligence (AI) for smart city scenario: Recent advancements and future trends. *Sensors* **2023**, *23*, 5206. [CrossRef]
42. Kashef, M.; Visvizi, A.; Troisi, O. Smart city as a smart service system: Human-computer interaction and smart city surveillance systems. *Comput. Hum. Behav.* **2021**, *124*, 106923. [CrossRef]
43. Guo, Q.B.; Wang, Y.; Dong, X.B. Effects of smart city construction on energy saving and CO_2 emission reduction: Evidence from China. *Appl. Energy*. **2022**, *313*, 118879. [CrossRef]
44. Jinxiu, W.; Kun, D. Impact and mechanism analysis of smart city policy on urban innovation: Evidence from China. *Econ. Anal. Policy* **2021**, *73*, 574–587. [CrossRef]
45. Ji, M.; Jin, M.; Chen, L.; Liu, Y.; Tian, Y. Promoting urban innovation through smart cities: Evidence from a Quasi-Natural experiment in China. *Environ. Stud.* **2024**, *13*, 319. [CrossRef]

46. Despoina, F.; Effie, K.; Lichao, W. Are smart cities green? The role of environmental and digital policies for Eco-innovation in China. *World Dev.* **2023**, *165*, 106212. [CrossRef]
47. Yan, Z.; Sun, Z.; Shi, R.; Zhao, M. Smart city and green development: Empirical evidence from the perspective of green technological innovation. *Technol. Forecast. Soc. Chang.* **2023**, *191*, 122507. [CrossRef]
48. Chu, Z.; Cheng, M.W.; Yu, N.N. A smart city is a less polluted city. *Technol. Forecast. Soc. Chang.* **2021**, *172*, 121037. [CrossRef]
49. Hoang, A.T.; Nguyen, X.P. Integrating renewable sources into energy system for smart city as a sagacious strategy towards clean and sustainable process. *J. Clean. Prod.* **2021**, *305*, 127161. [CrossRef]
50. Clement, J.; Manjon, M.; Crutzen, N. Factors for collaboration amongst smart city stakeholders: A local government perspective. *Gov. Inf. Q.* **2022**, *39*, 101746. [CrossRef]
51. Chen, Z.; Sivaparthipan, C.B.; Muthu, B. IoT based smart and intelligent smart city energy optimization. *Sustain. Energy Technol. Assessments* **2022**, *49*, 101724. [CrossRef]
52. Salman, M.Y.; Hasar, H. Review on environmental aspects in smart city concept: Water, waste, air pollution and transportation smart applications using IoT techniques. *Sustain. Cities Soc.*. **2023**, *94*, 104567. [CrossRef]
53. Ji, T.T.; Chen, J.H.; Wei, H.H.; Su, Y.C. Towards people-centric smart city development: Investigating the citizens' preferences and perceptions about smart-city services in Taiwan. *Sustain. Cities Soc.* **2021**, *67*, 102691. [CrossRef]
54. Chen, K.; Li, Q.; Shoaib, M.; Ameer, W.; Jiang, T. Does improved digital governance in government promote natural resource management? Quasi-natural experiments based on smart city pilot. *Resour. Policy* **2024**, *90*, 104721. [CrossRef]
55. Sharifi, A.; Khavarian-garmsir, A.R.; Kummitha, R.K.R. Contributions of smart city solutions and technologies to resilience against the COVID-19 pandemic: A literature review. *Sustainability* **2021**, *13*, 8018. [CrossRef]
56. Lee, J.; Babcock, J.; Pham, T.S.; Bui, T.H.; Kang, M. Smart city as a social transition towards inclusive development through technology: A tale of four smart cities. *Int. J. Urban Sci.* **2023**, *27*, 75–100. [CrossRef]
57. Clement, J.; Ruysschaert, B.; Crutzen, N. Smart city strategies–A driver for the localization of the sustainable development goals? *Ecol. Econ.* **2023**, *213*, 107941. [CrossRef]
58. Filippo, M.; Francesca, M.; Andrea, B. Smart city as a hub for talent and innovative companies: Exploring the (dis) advantages of digital technology implementation in cities. *Technol. Forecast. Soc. Chang.* **2023**, *193*, 122636. [CrossRef]
59. Caird, S. City approaches to smart city evaluation and reporting: Case studies in the United Kingdom. *Urban Res. Pract.* **2018**, *11*, 159–179. [CrossRef]
60. Singh, U.; Upadhyay, S.P.; Uttam, S. Fractured smart cities: Missing links in India's smart city mission. *Environ. Plan. B Urban Anal. City Sci.* **2023**, *50*, 1790–1805. [CrossRef]
61. Shin, S.Y.; Kim, D.; Chun, S.A. Digital Divide in Advanced Smart City Innovations. *Sustainability* **2021**, *13*, 4076. [CrossRef]
62. Hao, Y.; Wu, Y.; Wu, H.; Ren, S. How do FDI and technical innovation affect environmental quality? Evidence from China. *Environ. Sci. Pollut. Res.* **2020**, *27*, 7835–7850. [CrossRef]
63. Matteo, M.; Alessandro, S.; Luigi, M.; Ramakrishna, S. Strategies for Improving the Sustainability of Data Centers via Energy Mix, Energy Conservation, and Circular Energy. *Sustainability* **2021**, *13*, 6114. [CrossRef]
64. Qinyuan, W.; Wencui, D. Social Capital, Environmental Knowledge, and Pro-Environmental Behavior. *Int. J. Environ. Res. Public Health* **2022**, *19*, 1443. [CrossRef]
65. Haque, A.B.; Bhushan, B.; Dhiman, G. Conceptualizing smart city applications: Requirements, architecture, security issues, and emerging trends. *Expert Syst.* **2022**, *39*, e12753. [CrossRef]
66. Ahmad, K.; Maabreh, M.; Ghaly, M.; Khan, K.; Qadir, J.; Al-Fuqaha, A. Developing future human-centered smart cities: Critical analysis of smart city security, Data management, and Ethical challenges. *Comput. Sci. Rev.* **2022**, *43*, 100452. [CrossRef]
67. Ismagilova, E.; Hughes, L.; Rana, N.P.; Dwivedi, Y.K. Security, Privacy and Risks Within Smart Cities: Literature Review and Development of a Smart City Interaction Framework. *Inf. Syst. Front. A J. Res. Innov.* **2020**, *24*, 21–22. [CrossRef]
68. Laufs, J.; Borrion, H.; Bradford, B. Security and the smart city: A systematic review. *Sustain. Cities Soc.* **2020**, *55*, 102023. [CrossRef]
69. Spicer, Z.; Goodman, N.; Wolfe, D.A. How 'smart' are smart cities? Resident attitudes towards smart city design. *Cities* **2023**, *141*, 104442. [CrossRef]
70. Sepasgozar, S.M.; Hawken, S.; Sargolzaei, S.; Foroozanfa, M. Implementing citizen centric technology in developing smart cities: A model for predicting the acceptance of urban technologies. *Technol. Forecast. Soc. Chang.* **2019**, *142*, 105–116. [CrossRef]
71. Bansak, K. Estimating causal moderation effects with randomized treatments and non-randomized moderators. *J. R. Stat. Soc. Ser. A-Stat. Soc.* **2021**, *184*, 65–86. [CrossRef]
72. Kolak, M.; Anselin, L. A spatial perspective on the econometrics of program evaluation. *Int. Reg. Sci. Rev.* **2019**, *43*, 128–153. [CrossRef]
73. Ferman, B. Inference in differences-in-differences: How much should we trust in independent clusters? *J. Appl. Econom.* **2023**, *38*, 358–369. [CrossRef]
74. LeSage, P.J.; Pace, K.R. Spatial econometric modeling of origin-destination flows. *J. Reg. Sci.* **2008**, *48*, 941–967. [CrossRef]
75. Beck, T.; Levine, R.; Levkov, A. Big Bad Banks? The Winners and Losers from Bank Deregulation in the United States. *J. Financ.* **2010**, *65*, 1637–1667. [CrossRef]

76. Tone, K.; Tsutsui, M. An epsilon-based measure of efficiency in DEA—A third pole of technical efficiency. *Eur. J. Oper. Res.* **2010**, *207*, 1554–1563. [CrossRef]
77. Lu, D. China's Growth Slowdown and Prospects for Becoming a High-Income Developed Economy. *Asian Econ. Pap.* **2017**, *16*, 89–113. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

smart cities

MDPI

Article

Decentralized Incident Reporting: Mobilizing Urban Communities with Blockchain

El-hacen Diallo [1], Rouwaida Abdallah [2], Mohammad Dib [3], Omar Dib [4,5,*]

[1] Université Paris-Saclay, CNRS, LISN, Rue Raimond Castaing Bâtiment 650, 91190 Gif-sur-Yvette, France; diallo@lisn.fr
[2] Université Paris Saclay, CEA, LIST, 91120 Palaiseau, France; rouwaida.abdallah@cea.fr
[3] Engie, 1 Pl. Samuel de Champlain, 92400 Courbevoie, France; mohammad.dib@engie.com
[4] Department of Computer Science, Wenzhou-Kean University, 88 Daxue Rd, Ouhai, Wenzhou 325060, China
[5] Department of Computer Science, Kean University, 1000 Morris Avenue, Union, NJ 07083, USA
* Correspondence: odib@kean.edu

Highlights:

What are the main findings?

- Development of a decentralized incident reporting system leveraging blockchain and IPFS for urban emergency settings.
- Empirical validation demonstrates the system's feasibility and performance in terms of throughput, latency, and cost.
- The proposed system ensures real-time notification and accurate incident management through blockchain's immutable records.
- An incentive model using blockchain-based tokens enhances user participation and system sustainability.

What is the implication of the main finding?

- Enhanced transparency and accountability in urban incident reporting.
- Improved effectiveness and sustainability of urban management through active user engagement.

Abstract: This paper introduces an innovative response to the pressing challenge of rapid and effective incident detection and management in urban settings. The proposed solution is a decentralized incident reporting system (IRS) harnessing blockchain technology and decentralized data storage systems. By empowering residents to report incidents, the proposed IRS enables seamless real-time monitoring and intervention by relevant departments. Built on a blockchain foundation, the proposed solution ensures immutability, transparency, security, and auditability, enhancing data resilience and comprehensive applicability. The proposed system leverages the InterPlanetary File System (IPFS) for the storage of incident proofs to manage the blockchain size effectively. Through the proposed IRS, transparency is upheld, enabling complete auditability of incident details and required interventions by citizens, societal bodies, and governmental bodies. Moreover, an incentive model is introduced to encourage active participation in incident reporting, thereby enhancing the system's overall effectiveness and long-term sustainability. The proposed IRS integrates mobile technology to facilitate user engagement and data submission, essential for urban emergency management. Empirical validation using the Quorum–Raft blockchain demonstrates the feasibility of the proposed approach in terms of system throughput, incident reporting delay, blockchain size, and deployment cost. Specifically, the system maintains a latency of under 15 s even at high transaction rates, can handle up to 200 incidents per second, and is cost-effective, with deployment estimates for 16 organizations over five years being under 1.99 million USD. The method involves extensive testing with simulated incidents and user interactions to ensure robustness and scalability, showcasing the system's potential for effective emergency management in urban environments.

Keywords: urban emergency management; decentralized incident detection; mobile technology integration; civic engagement and participation; blockchain-based transparency

1. Introduction

The rapid growth of urban areas and the increasing frequency of disasters have created a need for strong urban emergency management (UEM) systems [1]. Cities now face various challenges, including natural disasters, technological hazards, and public safety threats [2]. As urban populations expand and infrastructure becomes more complex, efficiently managing and responding to emergencies becomes more crucial than ever.

Effective UEM involves planning, coordinating, and implementing measures to prevent, mitigate, respond to, and recover from emergencies and disasters [3]. This field manages natural disasters, technological hazards, and human-caused incidents to protect lives, property, and the environment. It requires collaboration among government agencies, non-governmental organizations (NGOs), private sector entities, and the public [4]. Key components include risk assessment, disaster preparedness, emergency response, and recovery efforts [5]. The goal is to minimize the impact of emergencies through proactive planning, efficient resource allocation, and timely response actions.

Given UEM's critical role, swift and efficient detection and response to emergencies are more important than ever in our urbanized and interconnected world [6]. These incidents include damage to public property, natural disasters, and public safety issues. The rapid identification and management of incidents are now essential elements of modern governance and public safety [7].

Advanced technologies, such as mobile technology, geographic information systems (GIS), early warning systems, and real-time communication tools, present a transformative opportunity to enhance UEM capabilities. By integrating these technologies and fostering a culture of preparedness, UEM aims to build resilient communities capable of withstanding and recovering from adverse events. Specifically, these technologies empower citizens to become vigilant observers, actively contributing to real-time incident reporting [8]. This potential forms the foundation of a decentralized IRS [9], representing a paradigm shift from traditional centralized reporting frameworks.

Centralized reporting frameworks face several challenges that limit their effectiveness. Reports must pass through multiple bureaucratic layers before reaching decision-makers, causing delays in information flow and slower response times [10]. These systems often lack transparency and accountability, making it difficult to track response actions. They are also vulnerable to single points of failure; a compromised central server can render the entire system inoperative [11]. Furthermore, centralized systems struggle with scalability, often becoming overwhelmed by high report volumes during major incidents, leading to bottlenecks and delays. These limitations highlight the need for decentralized systems that use modern technology to improve speed, transparency, and community engagement.

The proposed decentralized IRS aims to significantly improve the speed and efficiency of response mechanisms [12]. It overcomes the limitations of traditional incident reporting by adopting a wider perspective. This approach includes adaptability to various situations and fosters a cooperative relationship between the community and the governance structure. It serves as a platform that enables citizens to actively contribute to the well-being of their urban environment [13].

The vision for this decentralized system goes beyond just reporting incidents. It aims to create a versatile platform that can adapt to a variety of scenarios. The goal is to encourage community engagement and promote proactive civic participation [14]. The system's core strength lies not only in its ability to quickly detect and respond to incidents but also in its potential to serve as a dynamic and inclusive platform for citizens' involvement in urban governance [15].

The effective tracking of actions and interventions by local authorities and government offices within the decentralized IRS is crucial. This feature supports a robust and transparent governance process, ensuring prompt responses to reported incidents and being thoroughly documented [16]. By maintaining a comprehensive record of interventions, the platform establishes a transparent trail, allowing citizens to evaluate how effectively their local authorities address incidents. This tracking mechanism is vital for accountable governance, enabling citizens and administrators to assess the responsiveness of public services in incident management [17]. Furthermore, recorded interventions provide valuable data for policymakers and urban planners. These records offer insights into incident response patterns, help identify trends, and pinpoint areas needing strategic improvements. The ability to monitor and analyze the actions of governmental entities fosters public trust. It supports evidence-based decision-making, enhancing urban safety and security through effective incident reporting [18].

However, achieving this vision is challenging. Implementing such a decentralized system requires addressing complexities related to data security, the authenticity of incident reports, ensuring users' (i.e., residents') privacy, and seamlessly integrating with existing emergency response infrastructures [19]. These challenges are significant, but incorporating cutting-edge technologies, particularly blockchain, offers a viable solution to overcome them [10].

Owing to its features, blockchain is a promising technology for enabling a seamless incident reporting system (IRS) [20]. Blockchain introduces transparency, accountability, automation, and efficiency into the incident-reporting ecosystem. Its inherent characteristics, including immutability and decentralization, address concerns about the authenticity of reports and residents' privacy. This integration promises a robust foundation for creating a transparent, accountable incident-reporting platform.

Blockchain is essential for decentralizing the incident reporting process in urban areas [21]. Unlike traditional centralized systems, which often fail to capture the dynamism and immediacy required for effective incident management, blockchain technology provides a decentralized architecture that empowers citizens to participate actively in the reporting ecosystem. Acting as a distributed ledger, it ensures that incident data are accessible and verifiable by a network of participants rather than confined to a central authority.

The immutability of blockchain records guarantees that once incident reports are validated and recorded, they remain secure and unmodifiable [22]. This feature contributes to the credibility of the reported incidents and fortifies the system's integrity. Moreover, the decentralized network structure of blockchain enhances the platform's resilience against potential disruptions, ensuring continued functionality even in the face of node failures or network issues.

Real-time incident reporting and monitoring are imperatives in navigating the challenges of the urban environment [23]. The blockchain-based platform is crucial for achieving this real-time responsiveness. By leveraging the capabilities of distributed ledger technology, incidents can be reported, validated, and acted upon swiftly. The platform transcends the constraints of traditional reporting mechanisms, providing a dynamic and adaptive system capable of addressing the diverse and evolving nature of incidents in urban areas.

Ensuring the authenticity and security of incident messages is critical, given the nature of the information being transmitted [24]. With its cryptographic foundations and decentralized structure, the blockchain establishes an immutable record of incident reports. This not only safeguards the integrity of the information but also mitigates the risks associated with unauthorized alterations or malicious tampering [25]. The platform, therefore, becomes a secure and auditable repository for incident-related data.

Tracking and monitoring interventions by various stakeholders involved in the incident reporting process necessitate a comprehensive and transparent system [20]. The blockchain's transparency and immutability are pivotal in creating a traceable record of actions taken by emergency services, government agencies, and other entities. This feature

enhances accountability, offering insights into the effectiveness and timeliness of responses and contributing to an environment of responsive and responsible governance.

Introducing an incentive mechanism facilitated by tokens stored and managed on the blockchain is crucial for encouraging stakeholder participation [26]. The platform's integration of blockchain-based tokens provides a transparent and tamper-proof method for rewarding positive behaviors and contributions. This motivational tool fosters a collaborative ecosystem where stakeholders are actively engaged in the incident reporting and resolution processes.

Beyond its technological functionalities, the platform is a catalyst for promoting civic engagement [27]. By decentralizing incident reporting and response mechanisms, the blockchain-based system transforms citizens into active contributors to the well-being of their urban environment. It encourages a sense of shared responsibility and empowers individuals to play a direct role in shaping the safety and functionality of the community.

This research, therefore, embarks on a journey to redefine the dynamics of incident reporting in our urbanized world. It is a call to harness the potential of mobile technology, decentralization, and blockchain to create a resilient, community-driven reporting system.

1.1. Contributions

The main contributions of the proposed work can be summarized as follows:

- This work introduces a decentralized system for incident reporting in urban areas, leveraging blockchain technology and the InterPlanetary File System (IPFS).
- The proposed system ensures transparency and accountability in incident management through the immutable nature of blockchain, which permanently records all incident reports and subsequent actions.
- The proposed system ensures real-time notification of relevant departments and offers an evidence-based framework, enabling a higher level of accuracy for making effective decisions.
- The paper presents an incentive model using blockchain-based tokens to reward residents for accurate and timely incident reporting, enhancing residents' participation and the system's sustainability.
- The feasibility of the proposed system is demonstrated through empirical validation using the Raft consensus protocol, highlighting its performance in terms of throughput, latency, and cost.

1.2. Paper Organization

Table 1 presents the notations used in the entire paper. The remainder of this paper is organized as follows: Section 2 reviews existing works, requirements, and main stages of building an IRS. Section 3 details the proposed framework. Section 4 presents the architecture of the proposed IRS and the protocol workflow, showcasing the complete life cycle of an incident in the proposed scheme. Section 5 discusses the experimentation and analysis conducted to demonstrate the system's feasibility. Section 6 compares the proposed IRS with existing traditional systems. Finally, Section 7 concludes the paper and discusses future work.

Table 1. List of abbreviations.

Abbreviation	Description
IRS	Incident Reporting System
UEM	Urban Emergency Management
IPFS	Interplanetary File System
DApp	Decentralized Application
ISP	Identity Service Provider

Table 1. *Cont.*

Abbreviation	Description
BIM	Blockchain Incident Management
RISC	Record Incident Smart Contract
CID	Content Identifier
RAFT	Reliable, Replicated, Redundant, And Fault-Tolerant
IoT	Internet of Things
GIS	Geographic Information Systems
GPS	Global Positioning System
API	Application Programming Interface

2. Literature Review

2.1. Existing Incident Reporting Systems (IRSs)

In the evolving landscape of incident reporting and management, many new solutions have emerged to meet the diverse needs of various industries.

The Safety-Reports Incident Reporting App prioritizes real-time recording and investigation of workplace incidents, including injury and illness reports [28]. It aims to reduce documentation time through mobile devices, ensuring due diligence via a simple investigation method and enabling data analysis for preventive action planning.

The Trackforce Valiant Security Incident Reporting Software [29] provides a comprehensive solution for security personnel to report incidents. It leverages technology to enhance traditional reporting mechanisms, including time-stamped video, photo, audio, and GPS location data alongside textual evidence. This approach streamlines the reporting process and ensures the creation of credible reports that can withstand legal scrutiny. The system is designed to adapt to the specific needs of various security operations, offering features like digital tracking of incidents, simplification of guard reporting, and production of litigation-resistant reports.

The Risk Assessor's free Incident Reporting Software [30] streamlines the process of reporting workplace incidents. It enables users to efficiently complete Near Miss, Incident, and Ill Health Reports, integrating features such as PDF report generation, which can be downloaded or emailed. This solution emphasizes ease of use and accessibility, offering online and mobile app options for report creation and management. It also features analytics to leverage incident data for safety improvements, allowing for live safety statistics and insights into incident types and trends.

In [31,32], the authors focus on the Critical Incident Reporting System (CIRS) within healthcare to enhance patient safety. They evaluate CIRS's effectiveness in identifying risks and improving clinical risk management. They emphasize the importance of a supportive error-reporting culture and management commitment to patient safety. The core idea of CIRS is reporting observed safety-related events so that they can be systematically analyzed and serve the reporting person and others in the healthcare system.

For more examples, refer to the article [33] that comprehensively surveys some incident reporting and management systems. It highlights the evolution of these systems with the advent of smartphones and web-based platforms, allowing citizens to report incidents such as accidents, domestic violence, and burglaries through various means.

2.2. Blockchain-Based Approaches

Blockchain technology, first conceptualized in 2008, revolutionized digital transactions by introducing the concept of a decentralized ledger [34]. This technology has evolved significantly, extending its applications beyond cryptocurrency [35]. The main properties of blockchain include its decentralized nature, immutability, transparency, and enhanced

security [36]. These attributes make blockchain an ideal solution for various sectors seeking robust data integrity and trust mechanisms without centralized control [37].

The expanded use of blockchain across diverse sectors such as finance, supply chain management, healthcare, and governance exemplifies its ability to adapt and innovate within various operational frameworks [38]. As blockchain continues to evolve, its application in specialized systems, particularly in incident reporting, showcases its adaptability and the expansive range of its utility.

In [39,40], the authors propose a system that addresses the current limitations of traditional incident reporting systems by ensuring data integrity, enhanced transparency, and guaranteed traceability of reports. The importance of blockchain in this context lies in its ability to create a decentralized, immutable ledger for incident reports, which ensures security and privacy and facilitates real-time information sharing among stakeholders. This approach incentivizes reporting through rewards and significantly enhances patient safety and healthcare quality by enabling a more reliable and effective incident reporting and learning ecosystem. While the authors presented theoretical concepts in their work, it was notably limited to health data, and it lacked experimental validation. This omission makes it challenging to assess the feasibility of their approach in practical settings.

The approach in [41] is the proposal of a blockchain-based framework for aviation incident reporting to enhance safety. Its decentralized and immutable nature suggests blockchain can significantly improve the efficiency and security of incident reporting in the aviation industry by ensuring that data remain unaltered and transparently accessible to all stakeholders. This approach aims to foster a safety culture and continuous improvement within the aviation sector. Despite the importance of the use case, this tailored framework for the aviation sector lacks comprehensive details regarding user privacy, data confidentiality, scalability, and blockchain data storage issues. Moreover, the absence of experiments makes it difficult to assess its practical feasibility.

In [16], the authors present a system called BISCUIT for reporting blockchain security incidents, emphasizing the importance of human observations in detecting security issues. BISCUIT uses blockchain to make the reporting process transparent, verifiable, and resistant to tampering. Unlike the healthcare and aviation incident reporting systems previously discussed, BISCUIT focuses on enhancing the security of blockchain networks by structuring and securing the incident reporting mechanism itself. This approach ensures the integrity of security incident data, leveraging blockchain's decentralized and immutable features to address the challenges in centralized systems. Although the study was centered on the potential of human observations to aid in detecting blockchain security incidents, it failed to address incentive mechanisms or provide performance details. Furthermore, the security of data and the architecture of smart contracts were not analyzed, leaving significant gaps in evaluating their approach.

2.3. IRS Requirements

Our study of existing Incident Reporting Systems, alongside exploring barriers and challenges to effective implementation, has enabled us to establish a comprehensive list of requirements that an ideal reporting system should meet.

- **Usability** emerges as a pivotal aspect, encompassing a user-friendly interface and an experience devoid of issues [33]. Users in [42] expressed concerns about the usability of incident reporting systems, noting that the process is perceived as overly time-consuming, with reporting training efforts deemed futile.
- **The efficiency** of an IRS is contingent upon its ability to facilitate quick report submission and response, thereby necessitating a seamless performance [33].
- **Transparency** fosters trust and accountability. Transparent mechanisms engender confidence in the system and afford stakeholders insights into incident management processes [31,33].
- **Anonymity** is critical, particularly in contexts where fear of punitive measures or legal repercussions stifles reporting. Systems should afford users the option of anonymity

to mitigate apprehensions and encourage candid reporting [33,42]. For instance, refs. [39,43] note that within the medical sector, additional challenges of the effectiveness of the IRS include the fear of punitive measures and legal and financial repercussions. Another study [44] on road accidents in India highlighted that approximately 74% of witnesses are reluctant to report incidents or assist victims due to the fear of legal complications and police harassment.

- **Clarity** refers to the imperative to delineate what constitutes reportable incidents, underscoring the importance of clear reporting guidelines. The authors in [45] discuss the need to know what to report in the context of incident reporting systems in healthcare. This work critiques the initial simplistic ambitions of incident reporting in healthcare, comparing the ideal scenario to the aviation industry's "orange wire" concept [46], which symbolizes system-wide learning and rapid action. It reflects on the failure of the healthcare sector to develop a similar effective social infrastructure for investigating, understanding, and improving based on incident reports. It addresses misconceptions like "Report it all", "More is better", and "Incidents measure safety", among others. It points to a fundamental misunderstanding or misapplication of incident reporting principles, such as overemphasizing data quantity over quality, misusing reporting as a performance measurement tool, and lacking effective feedback and learning mechanisms. Another example is from [47], where in the aviation industry, safety investigators express concerns regarding the issue of excessive reporting, which may drown out critical safety signals amid a flood of irrelevant information.
- **Incentivization** is essential for robust reporting. Financial incentives, coupled with a culture that valorizes reporting, can promote users' engagement [31,48]. Authors in [48] added that incident reporting is of little value and wastes time, and they advised that financial incentives for generated reports might be a greater motivation for reporting.
- **Feedback** mechanisms assume significance in fortifying the reporting loop. Effective feedback acknowledges the contributions of reporters [41] and instills confidence by demonstrating responsiveness to reported incidents [39]. Furthermore, the lack of feedback mechanisms exacerbates usability challenges within incident reporting systems. For instance, in [49], over 50% of providers expressed uncertainty regarding their reports' acknowledgment and impact. Establishing mechanisms for post-reporting follow-up could potentially incentivize greater engagement in reporting practices, thereby fostering a culture of accountability and continuous improvement.
- **The security** and **confidentiality** of data are non-negotiable imperatives in IRS design [39,40]. Robust security measures are imperative to safeguard sensitive information, thereby assuring users of data integrity and privacy [33,39].
- **Scalability** emerges as a critical consideration in designing and implementing an IRS. As noted by [39], achieving scalability presents a significant hurdle, particularly in expanding IRSs to encompass a broader network of healthcare facilities. The scalability challenge extends beyond mere technical limitations; it encompasses the capacity of IRSs to accommodate increasing volumes of incident reports while maintaining system performance and reliability.
- **The cost effectiveness** of an IRS warrants careful consideration. While reporting systems offer immense potential for enhancing incident management, their deployment entails significant financial outlays [42].

2.4. IRS Main Steps

Effective incident reporting systems encompass a structured approach with distinct yet interconnected steps. We can distinguish between four main steps in this process: report submission, validation and verification, incident processing, and evaluation and learning. Each stage is crucial in ensuring incidents are accurately reported, thoroughly investigated, and appropriately addressed to enhance organizational safety and mitigate risks.

1. **Report Submission:** In the initial phase of the incident reporting process, individuals or stakeholders directly involved in or witnesses to an incident submit detailed reports outlining the event.
2. **Validation and Verification:** Following report submission, the validation and verification stage involves assessing the accuracy, credibility, and completeness of the reported information by relevant departments.
3. **Incident Processing:** Once reports are validated and verified, they enter the processing stage, where they are systematically reviewed, categorized, and prioritized based on their severity and potential impact. This stage also includes resolving the reported problem or accident, where appropriate actions are taken to address the incident and prevent its recurrence.
4. **Evaluation and Learning:** The final stage in incident reporting involves evaluating the outcomes of incident management efforts and capturing lessons learned to enhance future incident response protocols. Additionally, this stage includes the evaluation of the reporter's contribution to the incident resolution process. This may involve providing ratings or feedback to the reporter, such as assigning star grades or tokens, to acknowledge their efforts and facilitate continuous improvement in incident reporting practices.

Notably, none of the existing approaches cover all four steps simultaneously. Our approach addresses this gap by providing a comprehensive framework that integrates all four essential steps into a cohesive incident reporting and management system.

2.5. IRS and Smart Cities

The concept of smart cities revolves around leveraging advanced technologies to create more efficient, sustainable, and livable urban areas [50]. Smart cities utilize various technologies, including the Internet of Things (IoT), big data analytics, and artificial intelligence. These technologies enhance urban living, from transportation and energy management to public safety and citizen engagement.

A decentralized IRS aligns perfectly with the goals of smart cities [51]. By enabling real-time, citizen-driven incident reporting, the IRS enhances public safety and empowers residents to actively participate in the governance and maintenance of their urban environment. This participatory approach not only improves the responsiveness and efficiency of emergency management but also fosters a sense of community and shared responsibility among citizens.

Moreover, integrating the IRS with other smart city technologies can lead to even greater benefits. For example, data collected through the IRS platform can be combined with IoT sensors and GIS data to provide a comprehensive view of urban incidents, enabling more effective and timely interventions [52]. Additionally, the transparency and accountability ensured by our blockchain-based system can help build public trust and promote a more open and collaborative relationship between citizens and city authorities.

By addressing the critical need for efficient incident management in urban areas, the IRS contributes to the broader objectives of smart cities, making them safer, more resilient, and more inclusive.

3. Proposed Framework

The proposed system revolutionizes how incidents are reported in cities. It leverages cutting-edge digital tools, including blockchain and IPFS storage, to facilitate secure and transparent reporting. Combining these advanced technologies enhances security, fosters trust, reduces vulnerabilities, and ensures resilient storage. This represents a significant advancement in creating reliable and efficient incident reporting systems (IRS) for urban areas. In this section, we delve into the key components of the proposed system, providing an in-depth exploration of the stakeholders involved, the decentralized infrastructure utilized, and the crucial communication channels that link all system components.

3.1. System Actors

Several actors are integral to the functionality and success of the proposed system. These actors, ranging from users to departments, each have unique roles and responsibilities that contribute to the overall effectiveness of the IRS. Figure 1 outlines the use cases for the main actors in our system, the user, and the department, which are then elaborated upon.

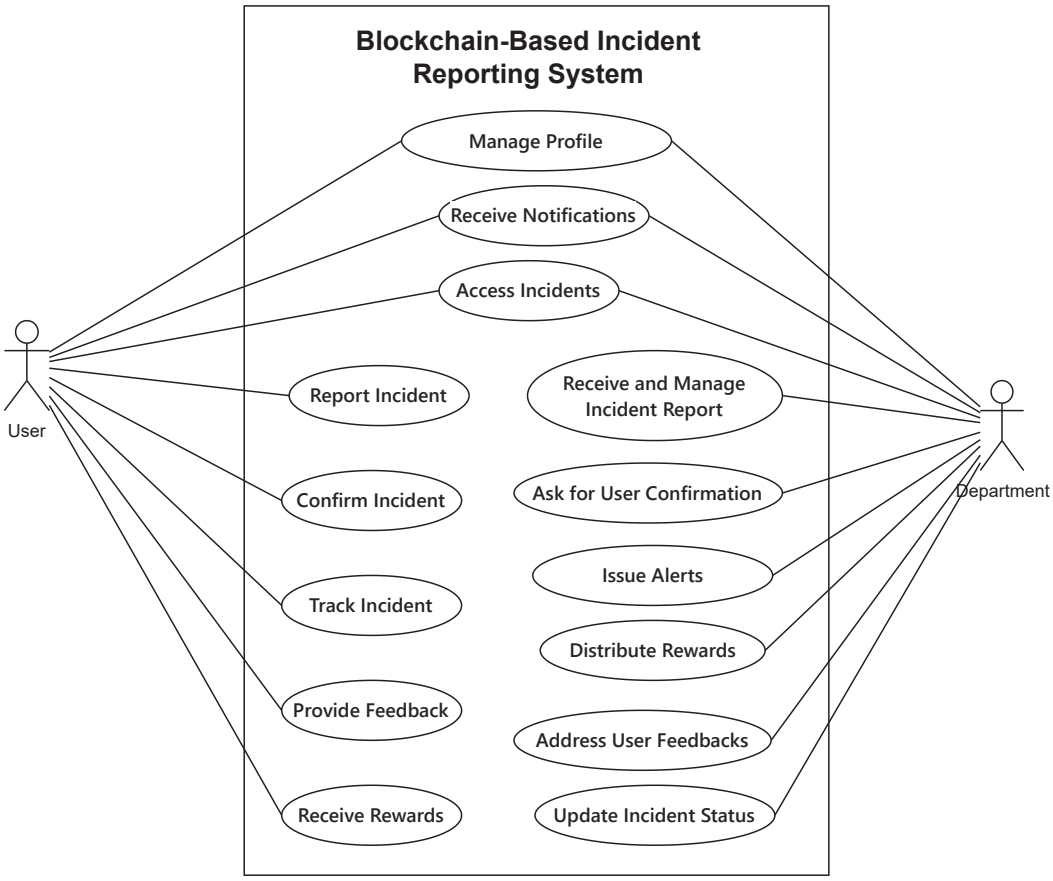


Figure 1. The use cases of the user and department actors.

3.1.1. User

The first actor in our system is the user. A user is a physical person who interacts with the system primarily through a dedicated mobile application. This mobile application serves as the main interface for accessing all features and functionalities related to incident reporting and management. The application contains the user interface and necessary APIs to communicate with the system.

Users can perform various actions, such as account management, reporting incidents, tracking incident status, confirming incidents, providing feedback, receiving rewards, accessing incidents, and receiving notifications. These actions are seamlessly facilitated through the user-friendly interface offered by a mobile application. In the following text, we detail these actions to provide a comprehensive overview of how they are carried out.

Account Management: Users commence their interaction by creating an account through the mobile application. Upon registration, they gain the ability to manage, update, and enhance their profile information directly within the application interface. This feature empowers users to modify personal details or add supplementary information, ensuring the accuracy and currency of their profiles.

Reporting Incidents: After completing their profile, users can initiate incident reporting via a dedicated and user-friendly interface after completing their profile. Through the mobile application, users can report various incidents, encompassing fires, road accidents, thefts, robberies, damages, and health emergencies, by furnishing comprehensive details

alongside evidence and proofs. They can augment their reports with descriptive text and multimedia evidence, such as photos, videos, or pertinent documents, to support their claims effectively.

Moreover, users can engage in the "witness" mode during event creation, converting the application into a smart witness. In this mode, the mobile application leverages specific phone sensors, including GPS location and humidity sensors, combined with AI-driven environmental metrics or other sensor data to gather additional evidence. Additionally, users can use in-app features for smart proof collection, which facilitates capturing environmental metrics or photos tailored to the nature of the reported incident.

These measures strengthen the accuracy and comprehensiveness of reported incidents, thereby assisting decision-makers and involved parties in evaluating incident quality and enhancing the responsiveness and reliability of the decision-making process.

Confirming Incidents: Upon initiating the creation of an incident, the user's input details undergo thorough cross-verification with existing active incidents stored in the system. This verification process aims to prevent the duplication of incidents within the system. If a match between the user's input and an existing active incident is found, the user is promptly notified and allowed to validate the matched event instead of creating a new one.

The validation process involves the user reviewing the details of the matched event and confirming its accuracy. This is facilitated through a dedicated validation report incident form, which closely resembles the incident creation form. Users can provide additional details, multimedia evidence, and sensor data to corroborate the validity of the matched event. By leveraging this validation process, the system ensures the accuracy and integrity of incident data while avoiding redundancy.

Tracking Incident Status: Furthermore, as the incident event lifecycle progresses, involved actors can track its status and evolution over time, enabling efficient management and resolution of reported incidents. This validation mechanism enhances the efficiency and reliability of the IRS, fostering trust and accountability among users and stakeholders.

For instance, upon detecting a new incident on the blockchain, involved departments can signify acknowledgment by adding a "received" flag to the incident record. Departments can then designate the incident as "under investigation", "confirmed", "resolved", "invalid", or other statuses to indicate its current stage in the resolution process.

The progress and actions of departments involved in incident resolution are subject to scrutiny by other entities and stakeholders in the system. Notably, any claims, status changes, or actions undertaken by departments are open to validation by the broader stakeholder community, reinforcing accountability and transparency throughout the resolution process.

Providing Feedback: Users can also provide feedback or comments specifically related to the resolution process or the outcomes of reported incidents. This feedback is tailored to the specifics of each incident and its handling. Examples of feedback may encompass various aspects, including the following:

- Delays: Users can highlight instances where delays in the resolution process impacted the timeframe for handling the incident.
- Inappropriate Handling: Users can flag instances where the event may have been mishandled or not appropriately addressed.
- Resource Insufficiency: Feedback may address cases where insufficient resources were allocated or utilized in addressing the reported incident.
- Output Quality: Users can comment on the quality or adequacy of the resolution's end result, ensuring that the outcome meets expected standards.

This feedback mechanism empowers users to contribute to continuously improving incident resolution processes. By providing insights into areas of improvement or commendation, users play an active role in enhancing the effectiveness and efficiency of incident handling within the system.

Receiving Rewards: Upon a user's creation of an incident and subsequent confirmation by a trusted official entity (e.g., fire department), rewards are distributed to various stakeholders involved in the incident reporting and confirmation processes. The distribution of rewards follows a predefined formula tailored to each stakeholder's role.

Creators and validators of incidents receive distinct allocations of rewards, reflecting their respective contributions. Early validators are incentivized with higher rewards to encourage prompt and accurate validation of incidents. Similarly, users providing accurate information or precise incident reports are rewarded at a higher rate to ensure that valuable contributions are duly recognized and incentivized.

Accessing Incidents Details: The proposed system enables users to request access to all incidents stored within it and selectively validate specific events, potentially earning rewards. This selective validation mechanism transforms each user into a potential witness, enhancing the reliability of stored information while incentivizing active engagement.

Users can view and access incident history and records, filtering incidents based on criteria such as timeframes, locations, or specific attributes. This streamlines the retrieval process according to user preferences. Additionally, users can perform queries for validation or statistical insights, examining details like incident frequency, types, and geographical distribution. These statistics provide valuable insights into patterns, trends, and concentrations over time and across locations. This feature allows users to monitor the progress and outcomes of incidents they have reported or engaged with.

Receiving Notifications: Users can receive notifications or alerts concerning updates on reported incidents, their resolutions, and the distribution of rewards. This feature ensures that users are kept informed about the progress and outcomes of incidents they are associated with. Additionally, users can opt to receive notifications or alerts prompting them to validate specific events. They can accept or decline these validation requests, streamlining the validation process by allowing users to prioritize their participation based on their availability and interest. Furthermore, users can receive real-time alerts tailored to incidents occurring within specified geographic areas, enabling prompt notification of incidents happening in locations of particular interest to them.

3.1.2. Department

The department stakeholder represents official entities responsible for executing specific tasks or providing services within a designated geographic area. These departments can be affiliated with governmental bodies or operate as standalone organizations. Examples include the Fire Department, Police Department, Identity Verification Department, Emergency Medical Services, City Government, Environmental Protection Agency, Public Health Department, Emergency Relief Organizations, Employment Verification Department, and Utility Companies.

Departments have a variety of responsibilities within the system, including incident management and response, updating incident status, asking for user confirmation, issuing alerts, providing feedback and continuous improvement, reward distribution, addressing user feedback, and accessing incident details. These functions are integral to ensuring the efficient and reliable operation of the IRS. In the following text, we detail these responsibilities to give an overview of their execution comprehensively.

Incident Management and Response: Departments oversee the operation and management of incidents within their jurisdiction. This includes monitoring and addressing various incidents such as fires, road accidents, thefts, robberies, damages, and health emergencies. Each department uses data from reported incidents to construct and execute its incident response strategies. For instance, the fire department uses these data to maintain a comprehensive overview of incident activity, ensuring timely response efforts.

Update Incident Status: Departments report and track real-time status updates and the progress of incidents, ensuring transparency and facilitating user audits. This provides stakeholders with up-to-date information on incident status, actions taken, and progress made, fostering accountability and trust within the IRS.

Ask for User Confirmation: Departments can request user confirmation for specific incidents to ensure the accuracy and validity of reported events. When a department identifies an incident that requires additional verification, it can prompt users who have reported or are located near the incident area to confirm its details. This process involves sending a notification to the relevant users, asking them to review the incident information and provide confirmation or additional evidence if needed. Users can respond through the mobile application to corroborate the incident's details, add multimedia evidence, or suggest corrections. This user confirmation mechanism helps departments verify the authenticity of incidents, reducing false reports and enhancing the reliability of the incident management system. By involving users in the confirmation process, departments also foster a sense of community participation and accountability.

User Engagement and Profile Building: Certain departments assist users in building their profiles within the system. For example, The Identity Verification Department verifies users' real identities and generates verifiable credentials. Other departments, such as the Employment Verification Department, provide credentials related to users' employment status and job titles. This support helps users develop the necessary skill sets and experiences for effective incident intervention. Some of these organizations play a crucial role in assisting users with building their profiles within the proposed system. For instance, the identity verification department aids in verifying users' real identities and generating verifiable credentials that may be pertinent in certain scenarios, such as proving residency within a specific area. Additionally, other departments are instrumental in helping users acquire verifiable credentials to streamline their involvement in specific incident circumstances. For example, the employment verification department may be tasked with providing users with verifiable credentials related to their employment status, job title, and the nature of their work. Furthermore, various departments may intervene at this stage to assist users in developing their skill sets and experiences, which are essential elements for effective intervention when necessary.

Issuing Alerts: A department can issue validation alerts within a designated area. For instance, the police department may send targeted alerts for specific incidents requiring validation, which is a relevant mechanism to mitigate false alerts and reduce intervention costs. This targeted approach ensures that validation efforts are focused on incidents requiring attention, optimizing resource allocation, and enhancing the efficiency of incident management processes.

Feedback and Continuous Improvement: Departments receive and address user feedback regarding reported incidents' resolution process and outcomes. This feedback mechanism allows users to highlight delays, inappropriate handling, resource insufficiency, and output quality. By incorporating user feedback, departments can continuously improve their incident resolution processes, ensuring effective and efficient handling of incidents.

Reward Distribution: Departments establish and manage reward mechanisms for incident reporting and validation stakeholders. Each department implements its system, including engagement tokens that can be exchanged for real money or used within the system for discounts and services. These reward systems are tailored to each department's specific needs and contexts, influenced by factors such as the nature of the incident, the availability of validators, and the geographical context. The transparency of these mechanisms enhances trust and credibility within the system, ensuring that stakeholders can track the distribution process. Additionally, departments can update their reward mechanisms based on user feedback and the behavior of validators, allowing for a dynamic and responsive approach to incentivizing participation and maintaining system integrity.

Similar to users, departments **have wallets** and can **access incidents**, **manage profiles**, and **receive notifications**. This allows them to filter, monitor, and analyze incidents for efficient response and strategic planning.

3.2. System Architecture

Integrating advanced technologies into the system architecture enhances its capabilities and functionalities, ensuring efficient incident reporting and management processes. This section delves into how various components work together seamlessly to achieve these goals. From the DApp acting as the central interface to using Smart Contracts and IPFS for secure and transparent transactions and data storage, each element is vital in fostering transparency, efficiency, and accountability within the ecosystem. Furthermore, the section explores the role of external tools, such as Decision Making Tools, Statistical Tools, and AI Tools, which augment the system's capabilities, empowering stakeholders with enhanced insights and functionalities. This comprehensive approach ensures that the system remains adaptable, responsive, and equipped to effectively address the evolving needs of incident resolution. Figure 2 illustrates the key components involved in the proposed reporting system.

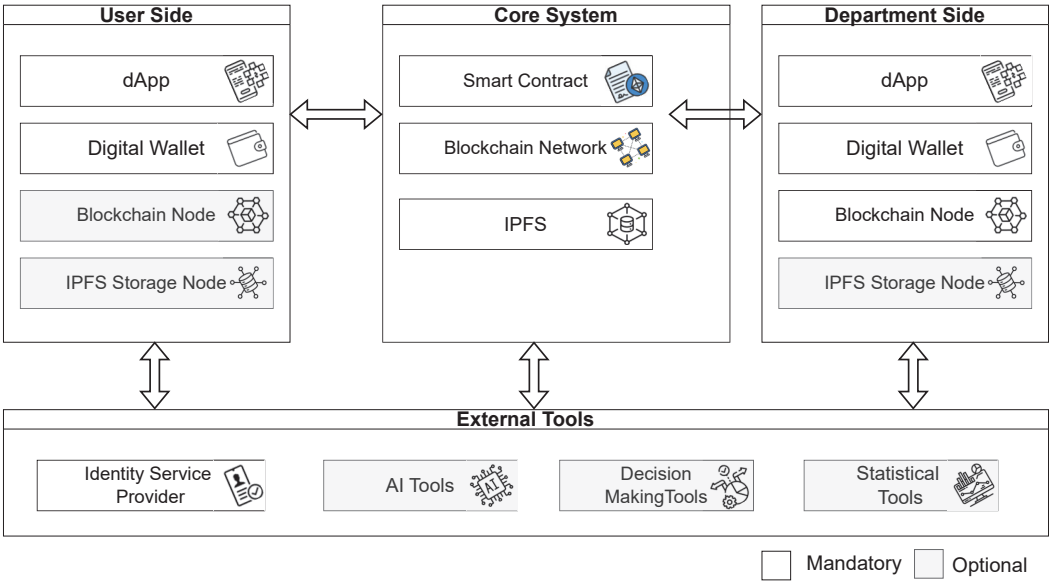


Figure 2. The technical components of the proposed system.

3.2.1. Decentralized Application (DApp)

DApp [53] acts as the system architecture's central interface for users and departments. It provides access to various features and functionalities related to incident reporting and management, catering to all stakeholders' needs.

The DApp offers users a user-friendly interface accessible through various platforms, including mobile devices and web browsers. It facilitates seamless user onboarding and identity verification processes, ensuring authenticity through verification by the Identity Service Provider (ISP). Additionally, users can communicate in real-time with other system actors, such as official departments and stakeholders, receive alerts, and participate in feedback and improvement processes. Prioritizing user experience and accessibility, the DApp empowers users to contribute effectively to incident reporting and management.

In contrast, for departments, the DApp provides essential tools and functionalities for efficient incident management, accessible through web-based interfaces or dedicated departmental systems. Departmental users can access features tailored to their roles, such as incident review, status updates, and reward distribution mechanisms. They can interact with incident reports, validate incidents, and coordinate response efforts seamlessly within the application. Moreover, the DApp enables departments to communicate with other stakeholders, receive alerts, and engage in feedback and improvement processes, fostering collaboration and enhancing the overall effectiveness of incident resolution.

3.2.2. Identity Service Provider (ISP)

The ISP plays a pivotal role in the system architecture, albeit external to it, by facilitating secure identity verification processes for users [54]. Entering personal details, which an official ISP verifies, initiates the user interaction with the system. To maintain user privacy, this sensitive information is not stored on the blockchain ledger. Instead, users engage with an official ISP (e.g., the state) through the dedicated mobile application to verify their information and link it to their digital identity within our system. This process ensures the privacy-preserving management of user information as identity details are transformed into verifiable credentials. Furthermore, official departments such as municipal offices or identity verification agencies can corroborate the user's information and issue verifiable credentials based on the user's provided data. These credentials are securely stored either locally on the user's device or in their personal cloud. Additionally, users can obtain verifiable credentials related to their profession or expertise by communicating with other official entities within the system. These credentials, too, are stored securely and utilized during incident reporting or intervention processes. Trusted official entities retain the authority to disclose user identity information if necessary, but only after confirming the user's misconduct. This dual mechanism ensures user information remains private and under their control while allowing official entities the capacity to disclose identity information or revoke verifiable credentials as needed, thus maintaining a balance between privacy and accountability.

For secure user authentication, a decentralized identity management system is essential. This approach should ensure robust authentication while also allowing users to report incidents confidentially. Currently, there is significant interest in proposing self-sovereign identities within blockchain technology to ensure user privacy and minimize proof disclosure [55,56]. For the proposed Incident Reporting System (IRS), we refer to the use of a self-sovereign identity scheme based on Schnorr Zero-Knowledge Proof (ZKP) to ensure user privacy as discussed in [57]. According to this protocol, a user can generate proof of being a citizen of a region and having the right to disseminate incidents without revealing their identity. The proof issuer, in this case, would be the state, which legally has the authority to verify citizen identities.

3.2.3. Digital Wallet

The Digital Wallet serves as a cornerstone of the user experience within the system, providing a secure and versatile platform for managing rewards and engaging with system features [58]. Each user account is seamlessly linked to a public wallet residing on the blockchain, facilitating the receipt of rewards as tokens. Users earn tokens for contributing valid incident reports or actively participating within the system. These tokens are securely stored within the user's wallet, serving as an incentive mechanism for sustained engagement and valuable contributions.

Similarly, departments within the system manage their reward mechanisms and have their own digital wallets tailored to their specific needs and contexts. Departments establish and manage reward systems for stakeholders involved in incident reporting and validation, implementing systems that may include engagement tokens exchangeable for real money or used within the system for discounts and services. These reward mechanisms are influenced by factors such as the nature of the incident, the availability of validators, and geographical context.

Moreover, the user's wallet boasts additional features to enhance user engagement and incentivization. These may include reputation metrics, reflecting the user's reliability and trustworthiness based on their contributions and interactions within the system. Community ranks, gamification-related indicators, and badges are also integrated into the wallet, giving users tangible recognition for their contributions and encouraging further engagement.

This hybrid storage approach ensures the security and privacy of user data. Verifiable credentials, such as identity information and user-generated data, are securely stored in

the user's digital wallet on their device. These credentials are shared only upon explicit user consent. Public information indicating the user's behavior within the system is stored on the blockchain, fostering transparency and accountability while safeguarding user privacy. This combination of local and public storage mechanisms contributes to a more trustworthy and secure system architecture, enhancing user confidence in their interactions with the platform.

3.2.4. Blockchain

The blockchain is the backbone of the IRS system architecture, facilitating transparent and secure record-keeping of all actions and interactions within the ecosystem. One of its primary functions is hosting the departments' and users' wallets and associated public data.

Additionally, the blockchain orchestrates the lifecycle of reported incidents and their progression, along with all related actions [59]. Actors within the system interact with the blockchain ledger to write, read, or update incident details, ensuring transparency and accountability throughout the incident management process. Furthermore, the blockchain manages the lifecycle of tokens within the user's wallet, ensuring seamless token management and secure transactions. Smart contracts facilitate and secure transactions, such as when a department transfers tokens to reward a user for their engagement.

To optimize efficiency and scalability, heavy incident information such as photos or videos is not directly stored on the blockchain. Instead, a decentralized file management system is employed, with unique identifiers for each resource stored on the blockchain. This approach ensures efficient incident management while preserving the integrity and scalability of the blockchain infrastructure.

The proposed system uses **Raft** [60] as a consensus mechanism to orchestrate the transaction lifecycle within a collaborative and distributed environment. This is reputable for its efficiency, reliability, and simplicity. Raft's leader election and log replication model ensure quick and consistent updates across the network, which is vital for maintaining the integrity of the system's data as it evolves through the project lifecycle.

Furthermore, Raft's ability to handle node failures gracefully ensures that the system's data remain available and consistent, even in the face of network or hardware failures (up to $((N-1)/2)$ failures, where N is the total number of participants in the consensus). This supports the high availability requirements of Blockchain Incident Management (BIM) applications. The immediate finality of data transactions under Raft, without the possibility of forks, guarantees that once a change is made to the BIM model, it is accepted and propagated throughout the network, ensuring all stakeholders are working from the same version.

This coherence is essential for avoiding costly misunderstandings and rework. In essence, Raft provides a robust, efficient, and straightforward consensus framework that aligns perfectly with the needs of BIM orchestration, fostering collaboration, enhancing data integrity, and supporting the real-time decision-making processes that are pivotal in modern construction and infrastructure development projects.

3.2.5. Smart Contracts

Smart contracts constitute another cornerstone of the system, residing on the blockchain to automate and enforce transactional agreements and workflows [61]. These self-executing contracts are programmable and transparent, ensuring the trustless execution of various processes within the ecosystem. Through smart contracts, users and departments can engage in secure and transparent transactions, with predefined rules and conditions enforced autonomously.

Moreover, smart contracts govern the lifecycle of incidents within the system, from creation to resolution, automating key stages of incident management. They enable the seamless progression of incidents through predefined workflows, ensuring adherence to established protocols and guidelines. Additionally, smart contracts can integrate external

data sources and oracles, enabling the system to access real-world information and trigger actions based on predefined criteria, enhancing responsiveness and effectiveness.

3.2.6. InterPlanetary File System (IPFS)

To alleviate the increasing size of the blockchain ledger and improve storage efficiency, our system leverages the IPFS [62] for managing and storing incident evidence. IPFS is a decentralized file storage and sharing network that allows for efficient, secure, and resilient data management by distributing it across multiple nodes. Unlike traditional centralized systems, IPFS addresses content by its unique hash (Content Identifier, or CID), ensuring data integrity and immutability.

Storing evidence on IPFS facilitates immediate access for emergency responses and serves as a permanent, auditable record of the incident and its associated evidence. This can be crucial for post-incident analyses, legal proceedings, and improving future emergency response strategies. Furthermore, the privacy and security of the evidence can be controlled using encryption. Files can be encrypted before being uploaded to IPFS, with decryption keys shared only with authorized parties. This ensures that the evidence remains secure and private while it is readily available for authorized users. The following steps showcase how evidence is stored in IPFS.

1. **Uploading Evidence:** Detailed evidence supporting the incident (such as photos, videos, documents, or sensor data) is uploaded to IPFS. Each file uploaded to IPFS is content-addressed, which means it receives a unique hash based on its content (CID).
2. **Immutable Records:** Once uploaded, the evidence cannot be altered without changing the file's hash. This immutability ensures the integrity of the evidence, providing a tamper-proof record of the incident and the associated response.
3. **Efficient Access:** The hash of the uploaded evidence is included in the PubSub notification, allowing for emergency services to access and retrieve the evidence directly from the IPFS network. IPFS's distributed nature ensures that files can be accessed quickly without relying on a central server.

3.2.7. Blockchain Node

Within the system architecture, users and departments can maintain their blockchain nodes, which is crucial for the operation and integrity of the decentralized network.

For departments, hosting a blockchain node is integral to their responsibilities within the system. These entities oversee the operation of the blockchain network, hosting and maintaining a full blockchain node, executing the consensus protocol, and defining the governance and rules of the system in collaboration with users. This collaborative effort involves reaching a consensus on transactions, writing them to the ledger, and disseminating blockchain blocks across the network. Additionally, they are responsible for deploying necessary smart contracts to facilitate various functionalities within our system. Given their greater resources and vested interest in decentralizing the IRS, these organizations are well-equipped to handle the technical responsibilities of maintaining blockchain nodes. Discussions may be held and subsequent implementations made regarding node revocation procedures, actions to be taken in case of revocation, protocols for addressing malicious behavior by blockchain actors, procedures for admitting new nodes into the system, and methodologies for updating smart contracts, among other considerations. This governance framework ensures the orderly functioning and evolution of the system, fostering transparency, accountability, and user participation in decision-making processes.

On the other hand, while users can maintain a full blockchain node, it may not be the most cost-effective choice. Therefore, we propose that the technical responsibility of hosting blockchain nodes be primarily entrusted to departments or organizations. Users can opt to interact with the blockchain network indirectly through the DApp, which abstracts away the complexities of directly managing a node. This approach allows users to focus on contributing to the IRS without the burden of maintaining infrastructure. At the same

time, departments or organizations ensure the robustness and reliability of the underlying blockchain network.

3.2.8. IPFS Storage Node

Users and departments can operate their IPFS nodes, which serve as decentralized storage solutions, facilitating storing and retrieving multimedia content associated with reported incidents.

Operating an IPFS node for departments can enhance their access and control over data management. These entities may choose to host their own IPFS nodes to ensure the availability and integrity of incident-related multimedia content. By maintaining their IPFS storage infrastructure, departments or organizations can efficiently manage large volumes of data while adhering to decentralized principles.

Similarly, users also have the option to operate their own IPFS storage nodes, providing them with greater control over the storage and sharing of multimedia content associated with their reported incidents. However, operating an IPFS node may entail technical complexities and resource requirements that not all users may be willing or able to manage.

To effectively deploy a decentralized file storage system, we propose that local authorities or trusted organizations manage the deployment of IPFS nodes, ensuring a stable and scalable infrastructure for users to connect to without needing to run their own nodes. User-friendly mobile and web applications will facilitate seamless interaction with the IPFS network, simplifying the processes of uploading, retrieving, and managing incident-related multimedia content. For advanced users and developers, secure APIs can be provided to enable integration with third-party applications, fostering an ecosystem of tools that enhance the incident reporting process.

Therefore, while both users and departments or organizations can operate IPFS storage nodes, it is important to note that this choice is optional and may not be suitable for all stakeholders. Users and departments can leverage other entities' existing IPFS infrastructure or integrate alternative storage solutions into the system architecture. Ultimately, the decision to operate an IPFS storage node should be based on technical expertise, resource availability, and data management requirements.

3.2.9. External Tools

The system architecture allows for seamless integration with external tools, enabling enhanced functionality and capabilities for decision-making, analysis, and automation. External tools such as Decision Making, Statistical, and AI tools can be interfaced with the system to augment its capabilities and provide additional insights and functionalities.

- Decision-Making Tools: External decision-making tools can be integrated into the system's components (i.e., IPFS and Blockchain) to facilitate informed decision-making processes [63]. These tools may include decision support systems, expert systems, or simulation tools that assist users and departments in assessing and evaluating incident data, identifying trends, and making strategic decisions to improve incident resolution outcomes.
- Statistical Tools: Statistical analysis tools can be integrated to enable comprehensive data analysis and visualization. These tools empower users and departments to perform advanced statistical analyses, generate reports, and derive insights from incident data. By leveraging statistical tools, stakeholders can identify patterns, correlations, and anomalies in incident reports, leading to more informed decision-making and proactive incident management strategies [64].
- AI Tools: Artificial intelligence (AI) tools can enhance the system's capabilities by automating repetitive tasks, predicting incident trends, and identifying potential risks or opportunities. AI algorithms can analyze large volumes of incident data, detect patterns, and make recommendations for optimizing incident resolution processes. Additionally, AI-powered chatbots or virtual assistants can provide real-time support and assistance to users, improving the overall user experience and the system's efficiency.

By interfacing with external tools, the proposed system harnesses the power of advanced technologies to streamline incident reporting and management processes, enhance data analysis and decision-making capabilities, and ultimately improve the effectiveness and efficiency of the incident resolution ecosystem.

4. Protocol

This section outlines the interactions among various components for secure and efficient incident reporting and management. Figure 3 illustrates the interactions among components using a fire incident use case. As depicted, when a resident (i.e., user) identifies a fire incident, they collect evidence using their mobile phone and submit it to the blockchain to initiate an incident report. The incident report evolves over time on the blockchain as additional evidence is gathered by other users, who update the report's status on-chain. The details of the incident evidence are stored on IPFS and linked to the blockchain ledger. Local authorities, connected to the blockchain in real time, analyze both the blockchain and IPFS data to facilitate decision-making. Figure 4 further depicts the system's workflow using a sequence diagram. The main steps of the proposed IRS are divided as follows:

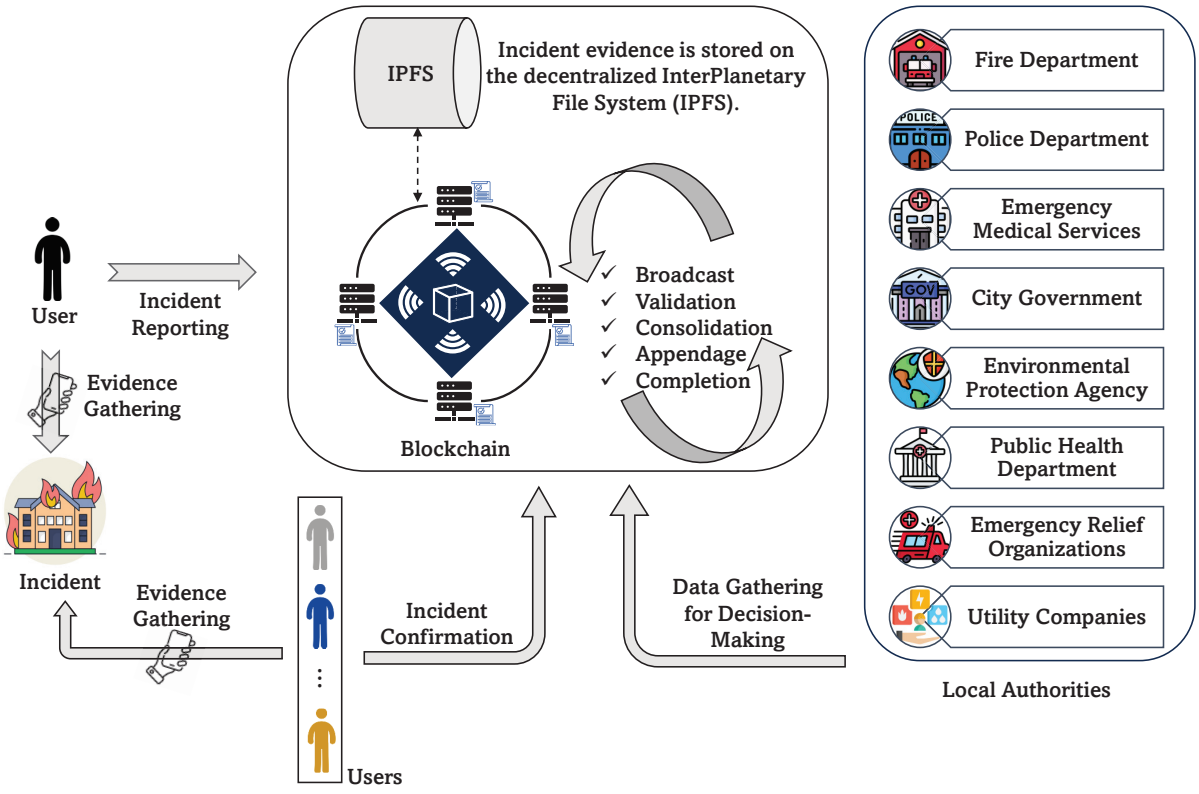


Figure 3. The architecture of the blockchain-based Incident Reporting System.

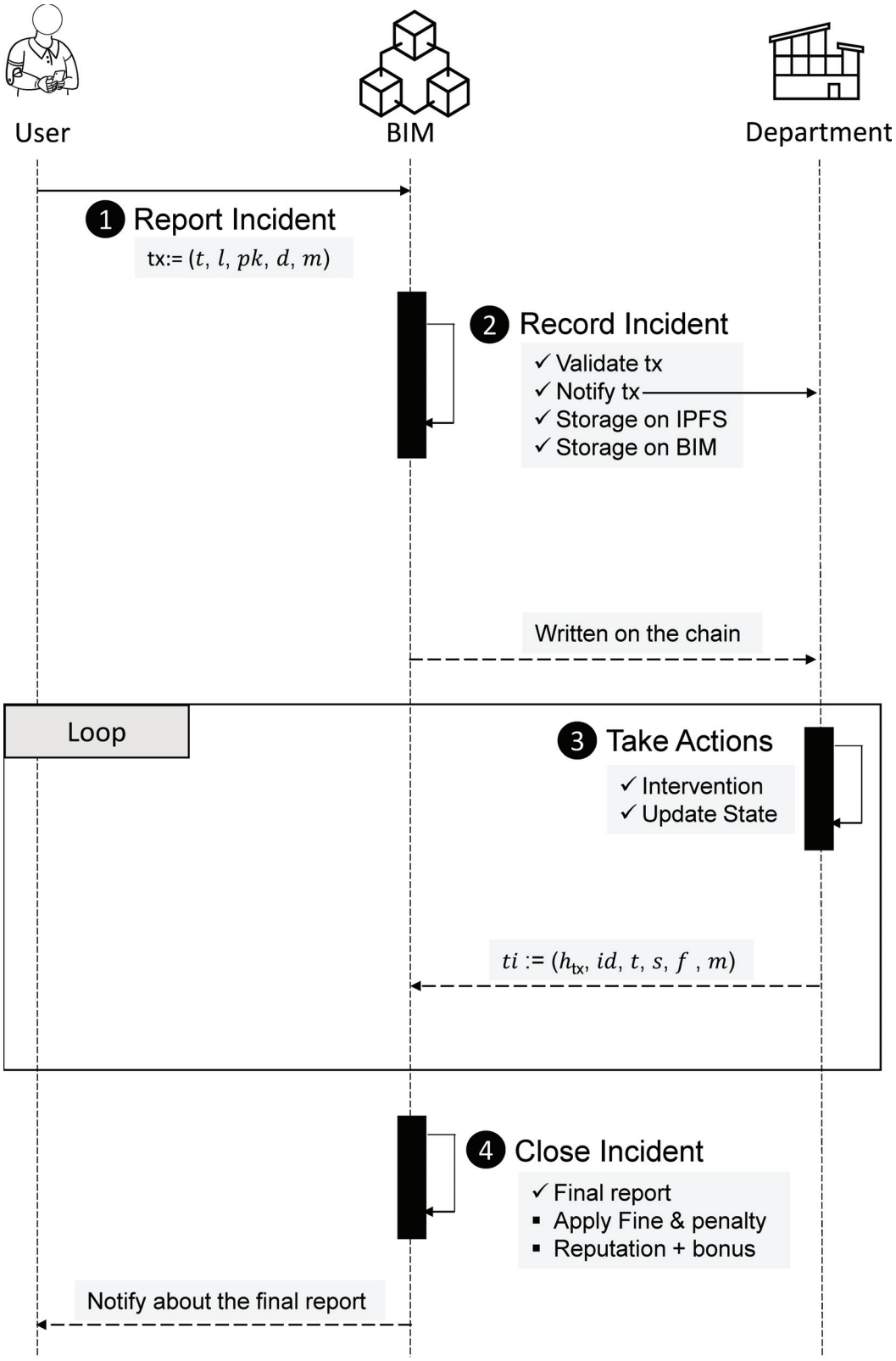


Figure 4. The workflow of the blockchain-based Incident Reporting System

4.1. Report Incident

Initially, users can seamlessly report an incident through a user-friendly application interface, equivalent to contacting the relevant emergency departments in the current sys-

tem. Upon activation, the incident is formalized as a blockchain transaction, encapsulating the essential details required for incident reporting and management:

$$tx = (t, l, pk, d, m),$$

where t is the timestamp, l is the location, pk is the public key of the user, d is the description of the incident, and m includes any attached media that offers additional proof or context for the incident.

This structure captures the fundamental details necessary for effective incident management.

4.2. Record Incident

When an incident report is verified and validated, the relevant departments are notified simultaneously when initiating IPFS storage. This parallel process helps save valuable seconds by alerting the departments about the incident, even if there is a delay in storing the transaction on the blockchain. Consequently, this ensures real-time notification of the incident to the relevant departments, even though it may take a few extra seconds for them to access the proof. This approach enhances the efficiency and effectiveness of emergency response efforts by ensuring timely alerts and prompt action. It is assumed that the emergency departments have dedicated infrastructure to monitor any updates on the blockchain, allowing them to instantly update the status of the incidents for which they are responsible.

The Record Incident Smart Contract (RISC) in Algorithm 1 illustrates the incident record logic. RISC incorporates transaction validation. If the reported incident includes media files, they are stored off-chain using IPFS. The IPFS storage is an asynchronous process. Once the media files are successfully stored on IPFS, the CID (Content Identifier) is included in the transaction metadata and recorded in the Blockchain for Incident Management (BIM). This ensures the link between the off-chain data and the on-chain transaction is secure and verifiable.

This approach preserves the integrity of the evidence by leveraging the strengths of both the blockchain and IPFS. When an incident is reported, any associated media files (such as photos or videos) are first uploaded to IPFS. IPFS, being a decentralized storage system, generates a unique Content Identifier (CID) for each file. This CID is then included in the incident report logged on the blockchain.

Recording the CID on the blockchain establishes a permanent and immutable link between the on-chain incident record and the off-chain media evidence stored on IPFS. This means the evidence uploaded to IPFS can be directly and securely linked to the immutable blockchain record. Including the CID in the blockchain ensures that any attempt to alter the media files on IPFS will result in a different CID, thereby breaking the link to the original incident record. This cryptographic linkage guarantees that the evidence cannot be tampered with without detection, thus maintaining the integrity and trustworthiness of the incident report.

Furthermore, storing large files directly on the blockchain is impractical and costly due to size limitations and high transaction fees. By using the blockchain to record incident details and IPFS to store the actual files, we optimize both storage efficiency and cost. The incident is recorded on the blockchain, including the hashes (CIDs) of the evidence stored on IPFS. This setup streamlines the retrieval process: stakeholders can quickly access the blockchain record, obtain the IPFS hashes, and retrieve the evidence from IPFS. This separation of responsibilities ensures the blockchain maintains a secure and immutable record, while IPFS provides efficient and scalable storage for large files.

Incidents must be recorded on the blockchain as soon as they are created. Given that an incident may be time-critical, the proposed system must ensure their real-time propagation to concerned services for immediate action, without extensive verification at this stage. Instead, the proposed protocol will rely on a post-validation model to penalize intentional false advertising or malicious behavior.

Algorithm 1 Record Incident Smart Contract (RISC)

Data Structure Incident (tx)
    t: timestamp
    l: location
    pk: issuer
    d: description
    m: c_{id} ▷ IPFS Content Identifier
    h: $hash$(tx) || c_{id}

procedure RECORDINCIDENT(l, d, c_{id})
    validate(pk) ▷ Ensure issuer is authenticated
    store(l, d, c_{id}) ▷ Store incident details
end procedure

function VALIDATE(issuer)
    if issuer is not registered or authenticated **then**
        return False
    else
        return True
    end if
end function

function STORE(l, d, c_{id})
    Create new tx with provided details
    Include h in tx metadata for off-chain evidence
    Save tx to blockchain
end function

Due to the need for real-time updates in managing advertising incidents and the legal implications of accessing such incidents, it is crucial to address the legal status of various organizations involved in incident management. Consequently, RAFT was chosen as a consensus mechanism for orchestrating BIM. Both governmental and non-governmental entities will maintain the blockchain. The latter can be entities that enforce transparency in incident reporting, representing the interests of users against errors or defiance by governmental organizations. This aspect highlights a compelling argument for using blockchain technology to enhance transparency, fairness, and automation in incident reporting. This added layer of transparency holds emergency centers accountable for any subpar service. Furthermore, the proposed architecture will encourage different emergency centers to strive for excellence in client service.

By initially recording the event on the blockchain, you create an immutable timestamped record of the incident. This step ensures that the event's occurrence is securely logged in a tamper-proof and verifiable way.

4.3. Take Actions

Once their format is validated, the relevant emergency center is notified about its respective incidents. This enables the handling of time-critical situations promptly. At this stage, the emergency center must act upon the reported incident. For example, in the case of a fire report, the fire department should proceed to the specified location to assess the situation. Subsequently, it must report the integrity of the event to the blockchain. Whether the incident is relevant or not, the emergency service will issue a transaction to the blockchain to confirm that an action has been taken, including available media evidence. The action taken is formalized as follows:

$$ti = (h_{\text{tx}}, pk, t, s, f, m),$$

ti includes a reference to the original event (h_{tx}), the intervention agent identity (pk), the intervention timestamp (t), the incident state after verification (s), which can be true or false, the monetary penalty if applicable (f), and finally, media files (m) may be included to support the claim for greater transparency. As detailed previously, the media evidence is stored on IPFS for auditability in case of inspection.

At the beginning of the intervention, such as the arrival of the first car or agent, a transaction (*ti*) is issued as proof that the emergency department has taken action. Additional (*ti*) are recorded on the BIM throughout the incident lifecycle to update its status, as shown in Figure 4.

4.4. Close the Incident

After the respective emergency department has taken action and it is assumed that no further updates regarding a particular incident are needed, a *ti* type transaction will be issued to close the incident. This transaction includes a complete report about the incident, which will be attached. The transaction will also include any penalties or rewards.

Indeed, to encourage user participation in the proposed IRS, an incentive system is implemented. This system rewards positive behavior and penalizes malicious actions. The reputation model is embedded within the Reputation Manager Smart Contract (RMSC), as shown in Algorithm 2.

RMSC retrieves each transaction along with the reporting user's current reputation, adjusting this reputation based on the relevance of the reported incident. It calculates a bonus using an exponential decay formula for relevant reports, incentivizing prompt submissions. The algorithm imposes fines for irrelevant or false reports, deducting a penalty from the user's reputation.

The updated reputation is then recorded within BIM. By incorporating a bonus system that exponentially decreases reporting delays and enforcing penalties for false reports, the incentive model maintains system integrity and motivates users to report responsibly.

Algorithm 2 Reputation Manager Smart Contract (RMSC)

Inputs:
$ti = (h_{\text{tx}}, id, t, s, f, m)$
MAX_B : Maximum possible bonus
α: Decay rate for the bonus
p: Fixed penalty on the user reputation

Output: updated user reputation

procedure UPDATEREPUTATIONANDFINE(ti)
    $tx \leftarrow$ getTx(h_{tx})
    $r \leftarrow$ getCurrentReputation($tx.pk$)
    if v is true **then** ▷ Incident is relevant
        $r \leftarrow$ CALCULATEBONUS($tx.t, ti.t$)
    else
        $r \leftarrow r - p$
        imposeFineOnUser($tx.px, ti.f$)
    end if
    updateReputationOnBIM(tx, r)
end procedure

function CALCULATEBONUS(t_b, t_e)
    $\Delta t \leftarrow t_e - t_b$ ▷ Time difference in seconds
    $bonus \leftarrow \text{MAX_B} \cdot e^{-\alpha \cdot \Delta t}$
    return *bonus*
end function

Such a model encourages a more vigilant and responsive community, where citizens are more likely to contribute valuable information for the collective benefit. This collaborative effort between the system and its users improves incident management efficiency. It fosters a culture of responsibility and accountability among citizens. By integrating this incentive mechanism, the RMSC effectively leverages the collective power of the community to enhance security and safety, making it a pivotal tool in the broader context of public emergency management and response systems.

5. Experimental Analysis

This section focuses on evaluating the feasibility of the proposed Incident Reporting System (IRS). It starts with the definition of the assessment metrics before diving into the simulation and presentation of the results.

5.1. Assessed Metrics

Below, we outline the assessed metrics to demonstrate the efficiency of the proposed IRS. These metrics encompass various aspects of system functionality and efficiency, including incident report/notify delay, the end-to-end delay from incident creation to treatment (i.e., concerned emergency services have taken action to confirm the incident veracity), blockchain size, cost-effectiveness, and scalability.

A **Incident Notification Delay:** An incident undergoes various processes throughout its lifecycle, with one pivotal stage being the notification of the relevant emergency. This delay is denoted as t_{notify}, where

$$t_{\text{notify}} = t_{\text{pub}} + t_{\text{ipfs}} + t_B \tag{1}$$

where:

- t_{pub} represents the delay between a user clicking "submit incident" and the incident being received on the blockchain network.
- t_{ipfs} is the delay in writing and reading the incident-supported proofs to the IPFS.
- t_B is the time required to record the incident on the blockchain; this time includes the transaction format validation.

B **Response Time:** This refers to the duration between initiating an incident and arriving at the first responding unit, such as a police car. This timeframe includes the notification and the time the concerned dedicated services take to take action. Therefore, the comprehensive response time for an incident can be quantified as follows:

$$t_{\text{response}} = t_{\text{notify}} + t_{\text{action}} + t_B \tag{2}$$

Here, t_{action} is the delay incurred by emergency services in responding to and acting upon the incident. This measure is crucial for evaluating the efficiency and effectiveness of the emergency response protocol. Here, t_B is the time delay for storing a ti transaction on the blockchain.

C **The throughput:** The throughput measures the number of incidents a system can process within a given time frame. It is expressed in incident per second (incident/s). The throughput is defined as

$$Th = \frac{N_{tx}}{t_B} \tag{3}$$

where N_{tx} is the number of pending incidents and t_B is the delay in writing to the Blockchain (BIM).

D **Blockchain Size:** In the proposed IRS, each reported incident initiates a sequence of blockchain storage operations to update the incident state. The first storage operation involves recording the incident's initial report on the blockchain before notifying the relevant emergency services. Subsequently, additional transactions ti are logged to record the response time and actions taken by the emergency services and the closing

report. During the period between the initial response and the closing report, multiple updates may be necessary to accurately reflect the state of the incident. Let ω denote the average number of updates required to maintain this accuracy. Consequently, the total size of the blockchain (S_B) can be expressed by the equation

$$S_B = S_{tx} + (S_{ti} \times \omega), \tag{4}$$

where S_{tx} represents the size of the transaction that logs the incident, and S_{ti} denotes the size of the transaction recording the action taken. $\omega \geq 1$ is the average number of updates needed to adjust the incident state.

E **Scalability:** The proposed system capability to accommodate an increasing volume of incident reports and a high arrival rate, as well as the growing number of involved organizations, is crucial. Scalability ensures that the system remains efficient and responsive as demand grows over time, allowing it to adapt to evolving requirements and maintain optimal functionality even under high usage and activity levels. Therefore, the evaluation considers different blockchain network participants (8 and 16 validators, i.e., the blockchain maintainers) and increasing incident arrival rates from 1 to 400 incidents per second.

F **Deployment Cost:** A dedicated consortium maintains BIM, so transaction fees are not directly applicable. The primary operational costs arise from maintaining and writing to the blockchain. Each organization must invest in a server capable of running a Quorum–Raft node. As the consensus is established through message exchanges, it does not require heavy computation or particular software, as in the case of Proof of Work (PoW). Additionally, costs may be associated with promoting the adoption of the proposed IRS scheme. For simplicity, we estimate only the costs of implementing the proposed framework. Therefore, we express the global cost as follows:

$$C_{\text{total}} = C_{\text{server}} \times N \tag{5}$$

where

- C_{server} represents the cost of acquiring and maintaining a server capable of running a Quorum–Raft node.
- N is the number of nodes in the network.

5.2. Results

The objective of this section is to demonstrate the feasibility of the proposed protocol rather than to present a complete implementation. Therefore, some findings from existing studies have been relied upon instead of conducting the implementation independently while showcasing the unique features of the proposed solution.

Table 2 summarizes various crucial variables within the system. Specific values are derived directly from their definitions, such as transaction size Table 3. In contrast, others are sourced from established benchmark studies, such as Quorum performance [65] and IPFS performance [66]. Furthermore, real-world data are utilized to ascertain the average time required for intervention. Specifically, statistics from the 2023 France Firefighters report delay and intervention delay serve as our reference [67].

Assuming a bandwidth of 100 Mbps, the transmission time for 1 MB of incident data is approximately 0.08 s. This brief communication delay is minimal due to the high-speed capabilities of modern internet infrastructure. Thus, for an average incident size (including media files) of 64 MB, $t_{\text{pub}} = 0.08 \times 64 = 5.12$ s.

Table 2. Summary of variables setting.

Variable	Description	Value Setting
N	Number of validators responsible for maintaining the blockchain network, executing the consensus, and managing smart contracts. Validators host full blockchain nodes and ensure the security and transparency of transactions by verifying and recording them on the blockchain ledger.	8, 16
t_{ipfs}	Delay taken to write and read the incident supported media proofs on IPFS	6.573 s for 64 MB [66]
t_{pub}	The delay between a user clicking "submit incident" and the incident being received by the blockchain network	5.12 s for 64 MB and 100 Mbps bandwidth
t_B	Time required to record the incident on the blockchain.	Quorum–Raft performance evaluation [68]
t_{Action}	Delay incurred by emergency services in responding to the incident.	16 min 49 s (average total intervention time in France for Firefighters) [67]
S_{tx}	Incident transaction size.	1071 Bytes
S_{ti}	Action feedback/update state transaction size.	264 Bytes

Table 3. Sizes of variables.

Variable	Representation	Size (bytes)
h	SHA-256 hash	32
t	Unix timestamp (int)	4
l	Coordinates (float)	16
pk	ECC public key	64
d	String	800 (upper bound)
m	IPFS CID	59
f	Amount	8
s	Boolean (true/false)	1
σ	ECDSA Signature	64
Size of $tx : (t, l, pk, d, m)$		$32 + 4 + 16 + 64 + 800 + 59 + 32 + 64 = 1071$ bytes
Size of $ti : (h_{\text{tx}}, pk, t, s, f, m)$,		$32 + 64 + 4 + 1 + 8 + 59 + 32 + 64 = 264$ bytes

Note: An additional 32+64 bytes for the transaction hash (h) and its signature ($sigma$) is included in each transaction's total size calculation, as each transaction must contain its hash and signature (σ).

5.2.1. Incident Notification Delay

The IRS significantly enhances the traditional emergency response process by reducing latency and increasing efficiency. The current system in France, with an average response time of 2 min 28 s for 6.3 incidents per second [67], involves multiple steps prone to delays and errors. In contrast, the proposed scheme maintains a latency of under 15 s, as shown in Figure 5, even at high transaction rates, enabling near-instantaneous processing of notifications. Using a smartphone application, citizens can report incidents directly, automatically providing crucial information such as location, incident type, and user credentials. This streamlines the response process, minimizes human error, and eliminates potential biases during verbal communication with emergency operators.

For example, in a car accident, a user can quickly send a notification with photos and precise GPS coordinates, ensuring that responders have accurate information immediately. This can save valuable time compared to describing the location and incident details over a phone call. Additionally, recording incidents transparently on the blockchain ensures full accountability and reduces the risk of discrimination.

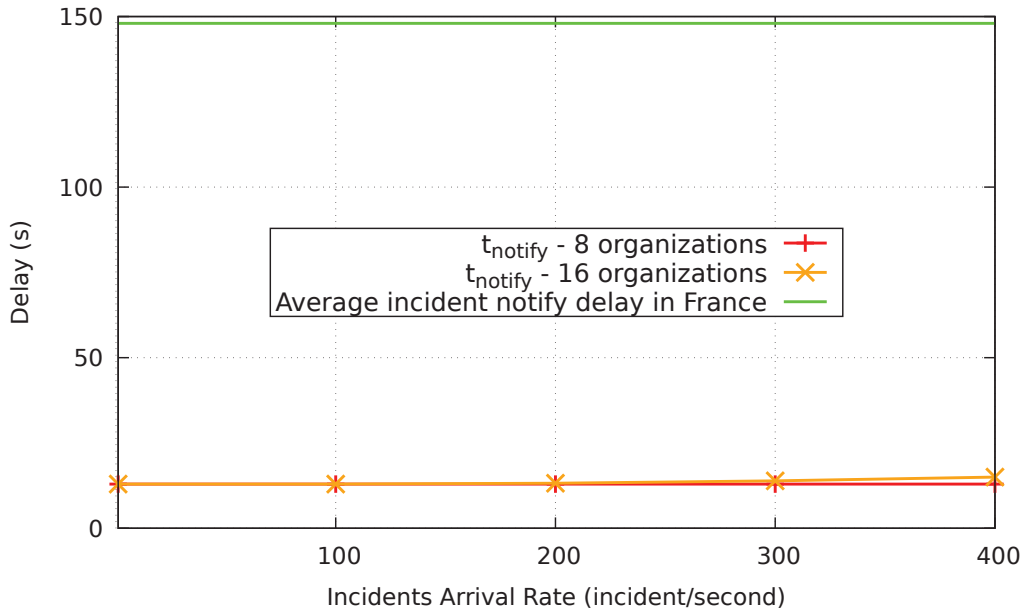


Figure 5. Incident notification delay.

Furthermore, the system supports governmental procedures by providing a reliable and tamper-proof record of all incidents. This transparency helps audit and analyze emergency responses, leading to better resource allocation and policy-making. While the integration of machine learning for automated incident assessment is a future perspective, the current advantages of the blockchain-based system in terms of speed, accuracy, and transparency make it a compelling alternative to the traditional emergency response approach.

5.2.2. Response Time

Figure 6 illustrates the required delay in response time, which is the time from incident issuance to the intervention of the concerned emergency services. The results show that the proposed IRS can reduce this delay compared to existing systems by more than 1 min (77 s). This improvement is primarily due to the reduction in the incident notification delay.

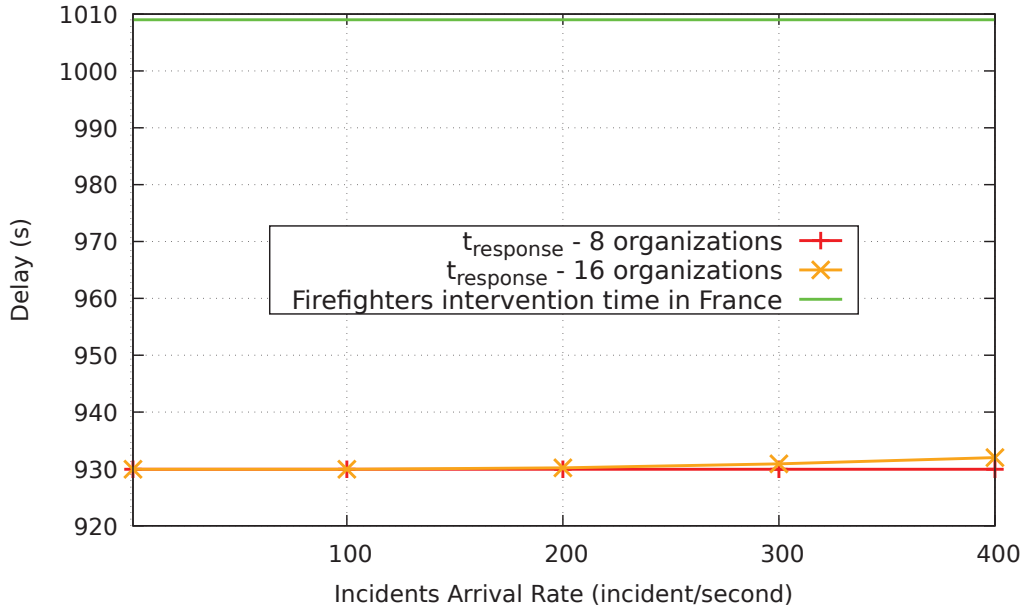


Figure 6. The Response Time.

The results also indicate that as the number of organizations maintaining the blockchain increases, the delay also increases. This is because of the increase in t_B, the time required to write on the blockchain.

Response time is an important metric for the IRS as it showcases the efficiency of the emergency services. Storing this information transparently on the blockchain helps to avoid complaints and makes the system more accountable.

5.2.3. Throughput

Figure 7 shows the number of incidents the BIM can validate and write on the blockchain against the incident arrival rate. The results indicate that at an incident arrival frequency of 200 incidents per second, approximately 50% of the transactions are instantly written on the blockchain, which is considered high.

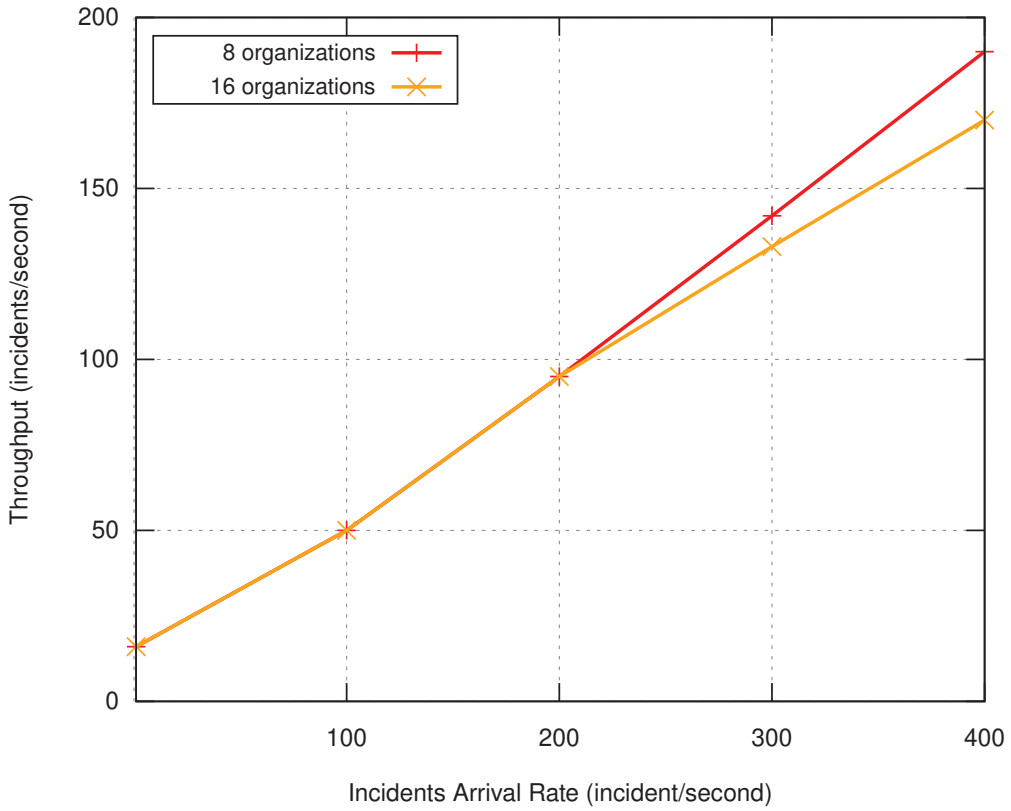


Figure 7. Throughput.

To put this result in context, in France, on average, 6.3 incidents were reported every second to the firefighters in 2023 [67]. Such a load can be handled without any bottleneck by the proposed IRS.

However, to anticipate the wider adoption of the proposed solution, considering different types of incidents beyond those concerning firefighters, we assess incident frequencies ranging from 100 to 400 incidents per second. This assessment ensures the IRS's scalability and efficiency in various real-world scenarios.

The results also demonstrate that the number of organizations involved impacts the throughput, adding further latency to the consensus process.

5.2.4. Blockchain Size

The proposed IRS enables comprehensive tracking of an incident lifecycle. This includes multiple updates to the status of incidents before they are finally closed. To objectively assess the size of the blockchain, Figure 8 plots the blockchain size with varying numbers of updates $\omega = [1, 2, 3, 4]$.

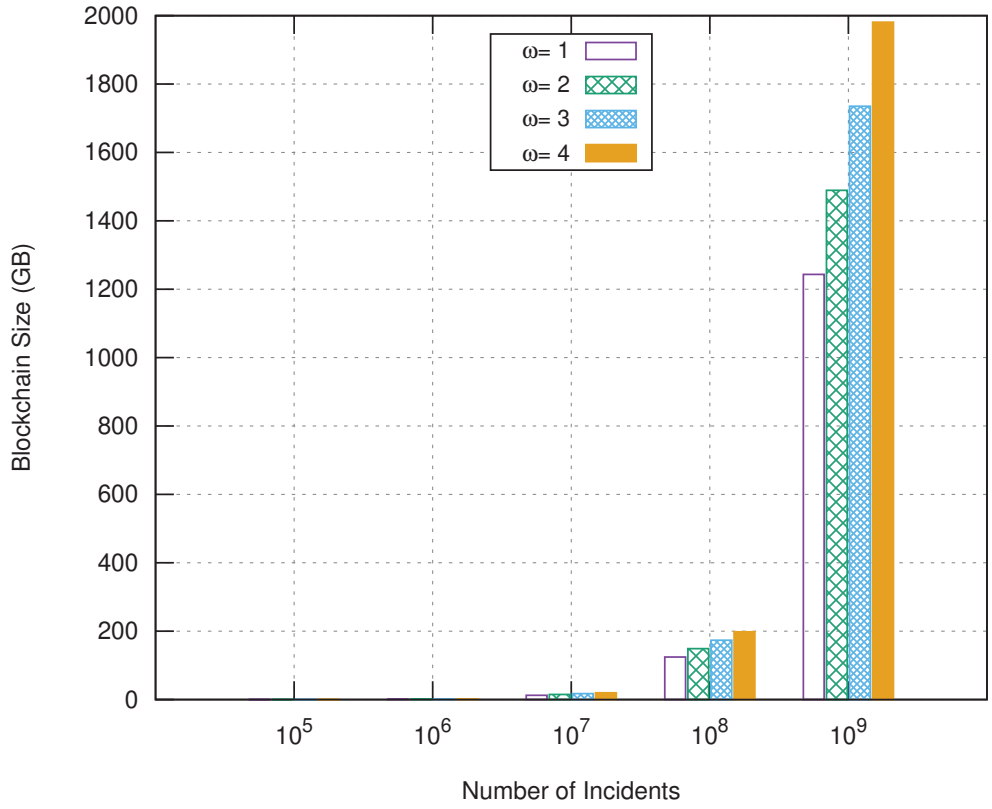


Figure 8. BIM size.

The results show that despite numerous updates to the blockchain ($\omega = 4$) to reflect the status of incidents, the blockchain size remains less than 200 GB with 10^8 incidents reported. This demonstrates the efficiency of the proposed IRS in maintaining a manageable blockchain size even with frequent status updates, owing to the off-chain storage using IPFS for media files supporting incidents.

To further elaborate, each incident can go through several stages, requiring updates to its status on the blockchain. These stages might include incident report, validation, response initiation, ongoing response updates, and incident closure. Even with multiple updates per incident, the proposed IRS ensures that the blockchain remains scalable and efficient in terms of storage.

Such a design is crucial for practical implementation, as it ensures that the blockchain does not become excessively large and unwieldy, which could compromise performance and accessibility. The ability to maintain a blockchain size under 200 GB with a large number of incidents highlights the robustness and scalability of the proposed IRS.

5.2.5. Deployment Cost

To accurately assess the costs associated with managing the proposed IRS system, we utilize the cost model of Hyperledger Fabric blockchain nodes on Amazon Managed Blockchain (Amazon Web Services, *Amazon Managed Blockchain Pricing*, [Online]. Available online: https://aws.amazon.com/managed-blockchain/pricing/ (accessed on 18 May 2024)). This choice is justified by the fact that Hyperledger Fabric, when running under the Raft consensus algorithm, has similar costs to a Quorum–Raft configuration. Table 4 breaks down the cost for Hyperledger Fabric on Amazon Managed Blockchain, and an estimation of C_{server}, the cost of maintaining a Quorum–Raft node.

Table 4. Cost components for Hyperledger Fabric on Amazon Managed Blockchain.

Cost Component	Description	Value (USD)/Month
Peer Node Monthly Cost (C_{peer})	Monthly cost per peer node	2014.07
Ordering Service Monthly Cost (C_{ordering})	Monthly cost for ordering service instance	487.91
Storage Cost (C_{storage})	Cost per GB of storage per month	0.10/GB

The total monthly cost C_{total} for managing N peer nodes, utilizing S_B GB of storage, is given by:

$$C_{\text{server}} = C_{\text{peer}} \times N + C_{\text{ordering}} + 0.10 \times S_B$$

$$C_{\text{server}} = 2014.07 \times N + 487.91 + 0.10 \times S_B$$

Thus, from the Equation (5), we have:

$$C_{\text{total}} = (2014.07 \times N + 487.91 + 0.10 \times S_B) \times N$$

Fixing the blockchain size S_B to 200 GB, Figure 9 plots the cost of maintaining 8 and 16 BIM nodes for varying numbers of years. The results show that over 5 years, it would cost less than 1.99 million USD to deploy the system with 16 organizations and 1 million USD for 8 organizations. This estimation provides a comprehensive evaluation of the cost of deploying the IRS.

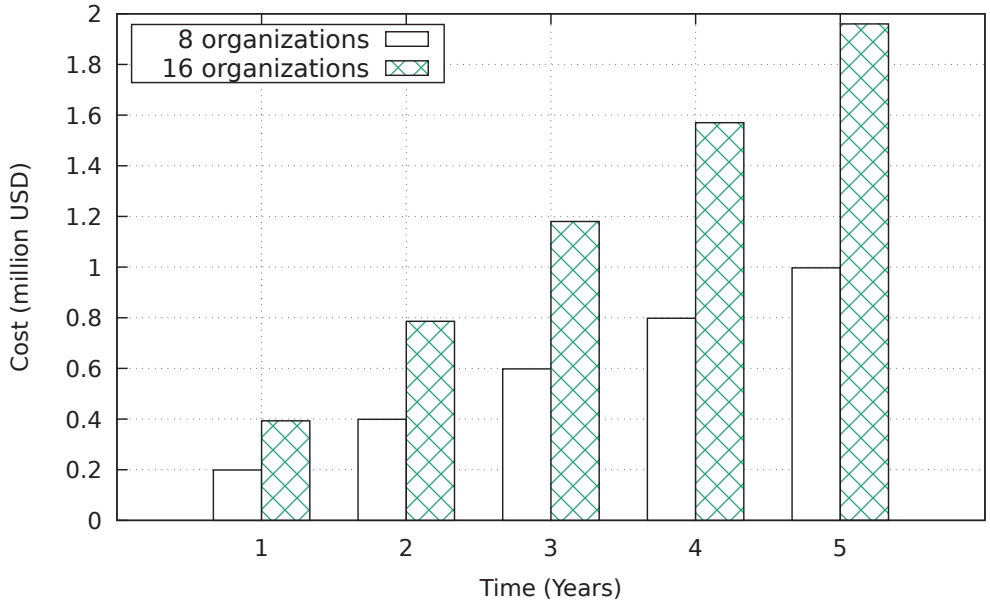


Figure 9. The total cost of deploying BIM.

While a significant budget would be needed to deploy the proposed IRS, the benefits include increased transparency, reduced errors, and enhanced collection of statistics, leading to better Quality of Service (QoS). Additionally, the potential integration of sophisticated AI for incident media analysis could further justify the deployment costs by significantly improving performance and efficiency.

6. Discussion

Comparing the proposed IRS with existing systems reveals several innovative enhancements that address key limitations in existing solutions.

Table 5 illustrates the benefits and limitations of the proposed approach compared to traditional and existing systems. The proposed decentralized IRS presents significant advancements over traditional centralized systems. By leveraging blockchain technology

and IPFS, the system ensures that all incident reports are transparent, immutable, and efficiently stored. This not only enhances the reliability of the data but also fosters a higher level of trust among users and stakeholders. The incorporation of a blockchain-based incentive model further motivates users to participate actively, thereby improving the overall effectiveness and responsiveness of incident management.

Additionally, unlike existing IRSs that often focus on isolated aspects like report submission or incident processing, the proposed IRS integrates all four critical steps: Report Submission, Validation and Verification, Incident Processing, and Evaluation and Learning. This comprehensive approach ensures accurate and thorough incident reporting while enhancing continuous improvement and accountability through effective learning and feedback mechanisms.

Furthermore, the IRS we propose also emphasizes user-centric features such as enhanced usability, incentivization, and feedback mechanisms, which are crucial for encouraging active engagement and participation. By integrating these elements, our proposed system offers a robust and scalable solution that overcomes the challenges of existing IRSs, thereby contributing valuable advancements to the field of incident management and smart city governance.

Table 5. Comparison between traditional systems and the proposed Incident Report System.

Feature	Traditional Systems	Proposed IRS
Transparency	Limited, controlled by central authority	High, due to immutable blockchain records
Accountability	Low, difficult to track and audit	High, all actions are recorded and traceable
User Anonymity	Generally low, users often disclose identity	High, allows anonymous reporting
Data Integrity	Vulnerable to tampering and loss	Strong, secured by blockchain and IPFS
Incentive Model	Typically none	Blockchain-based tokens to reward participation
Scalability	Established infrastructure, scales with investment	Potentially high, but requires robust infrastructure
Cost	Lower initial setup costs	Higher initial setup costs, potential long-term savings
User Engagement	Moderate, dependent on trust in the system	High, incentivized participation

7. Conclusions and Future Works

7.1. Conclusions

Integrating blockchain into urban incident reporting systems addresses critical requirements for effective incident management. These requirements include usability, efficiency, transparency, anonymity, incentivization, feedback mechanisms, data security, scalability, and cost-effectiveness. The proposed decentralized reporting system offers significant improvements over traditional centralized frameworks, meeting these requirements in several key ways.

Usability: The user-friendly mobile application enables citizens to easily report incidents, track their status, and receive updates. The system's intuitive interface ensures that users can participate in the reporting process without facing technical barriers, thereby enhancing civic engagement.

Efficiency: By utilizing blockchain technology in conjunction with the IPFS, the system ensures rapid incident reporting and data retrieval. The empirical validation based on the Raft consensus protocol demonstrates the system's capability to handle high throughput and low latency, crucial for timely incident management.

Transparency and Accountability: The immutable nature of blockchain ensures that all incident reports and subsequent actions are permanently recorded and verifiable. This transparency fosters users' trust and allows citizens and government bodies to audit the incident management process comprehensively.

Anonymity: The system supports a level of anonymity by allowing users to report incidents without directly revealing their identities on the blockchain. However, verifiable

credentials and identity verification through official entities ensure that user information can be corroborated if necessary. This balance addresses concerns about privacy while maintaining the integrity and trustworthiness of the system.

Incentivization: An integrated incentive model, facilitated by blockchain-based tokens, rewards users for accurate and timely incident reporting. This mechanism not only motivates active participation but also ensures the long-term sustainability of the system.

Feedback Mechanisms: The platform includes robust feedback loops where users can provide input on the resolution process. This continuous feedback helps improve the system and ensures that users feel their contributions are valued and impactful.

Data Security and Confidentiality: Blockchain's cryptographic foundations and the use of decentralized storage via IPFS ensure that incident data is secure, unalterable, and accessible only to authorized parties. This level of security is critical for maintaining the integrity and privacy of the reported incidents.

Scalability: The decentralized architecture and the use of IPFS for storing large files ensure that the system can scale to accommodate increasing volumes of incident reports and expanding networks of users and organizations. This scalability is essential for urban areas with growing populations and incident reporting needs.

Cost-Effectiveness: The cost of the proposed system has been analyzed based on available data from online servers, indicating an expense of 1 million USD over 5 years for a blockchain comprising 8 organizations. This cost is distributed among the participants, effectively decentralizing the financial burden.

Additionally, the proposed system comprehensively covers the four steps of the Incident Reporting System lifecycle:

Report Submission: Users can submit detailed incident reports through the mobile application, including descriptions and multimedia evidence.

Validation and Verification: The system validates the authenticity of reports using blockchain's immutable ledger and cross-verification with existing incidents.

Incident Processing: Once validated, incidents are processed and managed by relevant departments, ensuring timely and effective responses.

Evaluation and Learning: The system tracks and documents all interventions, allowing for continuous evaluation and improvement of incident management practices.

The proposed decentralized incident reporting system is designed to complement and enhance existing emergency response frameworks rather than replace them. One of the significant added values is the system's capability to empower citizens to report incidents seamlessly using their smartphones. This feature streamlines the reporting process, allowing users to easily document incidents such as accidents on an autoroute without the burden of calling emergency services and explaining the situation. The system holds users accountable for their reports through a verification process, ensuring that only accurate and legitimate incidents are recorded. This accountability not only deters false reporting but also enhances the reliability of the data received by emergency responders. Furthermore, the system incentivizes user engagement by offering rewards for accurate and timely reports, encouraging active participation in maintaining public safety. By enabling direct user participation, verification, and incentives, the proposed system reduces response times, improves the efficiency of incident management, and fosters a proactive community culture, all while building public trust through a transparent and accountable reporting process.

In conclusion, the proposed blockchain-based incident reporting system effectively meets the highlighted requirements, providing a robust, transparent, and user-friendly platform for urban emergency management. By addressing the critical needs of usability, performance, transparency, anonymity, incentivization, feedback mechanisms, data security, scalability, and cost-effectiveness, and by covering all four steps of the IRS lifecycle, the system represents a significant advancement in mobilizing urban communities for proactive incident reporting and management.

7.2. Future Works

This work proposes a decentralized framework for effective incident reporting and management in urban areas. Despite the detailed description, the following tasks are planned for future work:

The proposed framework includes a reputation model to track the reputation of actors based on their actions within the system. However, experimenting with this model has not yet been done. We aim to study how this model can punish bad actions and encourage good behavior within the system.

We also plan to investigate the IRS privacy issues thoroughly. For instance, sharing the location of an incident might inadvertently reveal the reporter's identity. To address this, we intend to integrate advanced cryptographic tools, enabling the sharing of sensitive information without compromising participant privacy. Another aspect of enhancing privacy is enabling data sharing between specific departments rather than placing all data on the blockchain, which is accessible to everyone. This can be achieved using the private channel features of the Quorum blockchain. Studying the feasibility and impact of this approach on system performance will be a valuable area of future research.

Another point involves automating more processes using AI techniques. For example, we aim to study how blockchain data can be used alongside AI tools to help departments predict the sensitivity of an incident based on a wide range of data and evidence.

Finally, implementing multiple reporting blockchains in neighboring cities is another consideration. Enabling these blockchains to interoperate and work together while respecting their own architecture and consensus mechanisms presents a significant challenge related to interoperability.

Author Contributions: Conceptualization, E.-h.D., R.A., M.D. and O.D.; methodology, E.-h.D., R.A., M.D. and O.D.; software, E.-h.D. and O.D.; validation, E.-h.D., R.A., M.D. and O.D.; investigation, E.-h.D., R.A., M.D. and O.D.; resources, E.-h.D., R.A., M.D. and O.D.; data curation, E.-h.D., R.A., M.D. and O.D.; writing—original draft preparation, E.-h.D., R.A., M.D. and O.D.; writing—review and editing, E.-h.D., R.A., M.D. and O.D.; visualization, E.-h.D., R.A., M.D. and O.D.; supervision, O.D.; project administration, O.D.; funding acquisition, O.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Wenzhou-Kean University (Supporting Project No. KY20231019000165).

Data Availability Statement: The original contributions presented in the study are included in the article/supplementary material, and further inquiries can be directed to the corresponding author.

Acknowledgments: The authors would like to express their sincere gratitude to the reviewers for their thorough and insightful comments. Their constructive feedback has been invaluable in refining our research and enhancing the overall quality of this paper. The authors also extend their gratitude to Wenzhou-Kean University for providing the necessary laboratory facilities and execution resources, which were instrumental in completing this work.

Conflicts of Interest: Author Mohammad Dib was employed by the company Engie. This affiliation did not influence the research outcomes. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Hernández, J.Z.; Serrano, J.M. Knowledge-based models for emergency management systems. *Expert Syst. Appl.* **2001**, *20*, 173–186. [CrossRef]
2. Elvas, L.B.; Mataloto, B.M.; Martins, A.L.; Ferreira, J.C. Disaster Management in Smart Cities. *Smart Cities* **2021**, *4*, 819–839. [CrossRef]
3. Liu, H.; Li, Y. Smart cities for emergency management. *Nature* **2020**, *578*, 515–516. [CrossRef]
4. Lee, D.W. The expertise of public officials and collaborative disaster management. *Int. J. Disaster Risk Reduct.* **2020**, *50*, 101711. [CrossRef]

5. Feng, Y.; Cui, S. A review of emergency response in disasters: Present and future perspectives. *Nat. Hazards* **2021**, *105*, 1109–1138. [CrossRef]
6. Costa, D.G.; Peixoto, J.P.J.; Jesus, T.C.; Portugal, P.; Vasques, F.; Rangel, E.; Peixoto, M. A Survey of Emergencies Management Systems in Smart Cities. *IEEE Access* **2022**, *10*, 61843–61872. [CrossRef]
7. FitzGerald, G.; Abrahams, J.; Pizzino, S. Principles of incident management. In *Disaster Health Management*; Routledge: London, UK, 2024; pp. 203–215.
8. Bustillo, J.C.M.; Mateo, J.T.I. Automated Incident Reporting Management System Using Mobile Technology. *Int. J. Innov. Manag. Technol.* **2020**, *11*, 18–26.
9. Michail, A. Tackling the Challenges of Information Security Incident Reporting: A Decentralized Approach. Ph.D. Thesis, University of East London, London, UK, 2020. [CrossRef]
10. Bhushan, B.; Khamparia, A.; Sagayam, K.M.; Sharma, S.K.; Ahad, M.A.; Debnath, N.C. Blockchain for smart cities: A review of architectures, integration trends and future research directions. *Sustain. Cities Soc.* **2020**, *61*, 102360. [CrossRef]
11. Biswas, K.; Muthukkumarasamy, V. Securing Smart Cities Using Blockchain Technology. In Proceedings of the 2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Sydney, NSW, Australia, 12–14 December 2016; pp. 1392–1393. [CrossRef]
12. Hu, W.; Chen, B.; Winter, S.; Khoshelham, K. Decentralized management of ephemeral traffic incidents. *Trans. GIS* **2022**, *26*, 2188–2205. [CrossRef]
13. Nazemi, A.R.; Dolatshahi, M.; Kerachian, R. A decentralized multi-agent framework for urban flood management. *Sustain. Cities Soc.* **2024**, *106*, 105328. [CrossRef]
14. Zengeya, T.; Mackenzie, R.; Nleya, S.M.; Dube, S.S.; Ndlovu, S.; Marabada, N.; Mutengeni, J.; Mapenduka, W. Incident Reporting System: A case for Bulawayo. In Proceedings of the 2023 2nd Zimbabwe Conference of Information and Communication Technologies (ZCICT), Gweru, Zimbabwe, 2–3 November 2023; pp. 1–5. [CrossRef]
15. Kaswan, K.S.; Gautam, R.; Dhatterwal, J.S. Introduction to DSS System for Smart Cities. In *Decision Support Systems for Smart City Applications*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2022; Chapter 4, pp. 53–76. [CrossRef]
16. Putz, B.; Vielberth, M.; Pernul, G. BISCUIT - Blockchain Security Incident Reporting based on Human Observations. In Proceedings of the 17th International Conference on Availability, Reliability and Security, Vienna, Austria, 23–26 August 2022. [CrossRef]
17. Merrell, I. Blockchain for decentralised rural development and governance. *Blockchain Res. Appl.* **2022**, *3*, 100086. [CrossRef]
18. Khan, Z.; Abbasi, A.G.; Pervez, Z. Blockchain and edge computing–based architecture for participatory smart city applications. *Concurr. Comput. Pract. Exp.* **2020**, *32*, e5566. e5566 CPE-18-1400.R2, [CrossRef]
19. Ridjic, O.; Jukic, T.; Ridjic, G.; Mangafic, J.; Busatlic, S.; Karamehic, J., Implementation of Blockchain Technologies in Smart Cities, Opportunities and Challenges. In *Blockchain Technologies for Sustainability*; Springer: Singapore, 2022; pp. 71–89. [CrossRef]
20. Makani, S.; Pittala, R.; Alsayed, E.; Aloqaily, M.; Jararweh, Y. A survey of blockchain applications in sustainable and smart cities. *Clust. Comput.* **2022**, *25*, 3915–3936. [CrossRef]
21. Wang, Y.; Chen, H. Blockchain: A potential technology to improve the performance of collaborative emergency management with multi-agent participation. *Int. J. Disaster Risk Reduct.* **2022**, *72*, 102867. [CrossRef]
22. Jlil, M.; Jouti, K.; Boumhidi, J.; Loqman, C. Blockchain-Based Mobile Application for Traffic Accident Report's Management. In Proceedings of the 2024 4th International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), FEZ, Morocco, 16–17 May 2024; pp. 1–6. [CrossRef]
23. Jiang, X.; Shi, Q.; Miao, H.; Cao, W.; He, H.; Chen, S.; Yang, J. Credible Link Flooding Attack Detection and Mitigation: A Blockchain-Based Approach. *IEEE Trans. Netw. Serv. Manag.* **2024**, *21*, 3537–3554. [CrossRef]
24. Chinnasamy, P.; Vinothini, C.; Arun Kumar, S.; Allwyn Sundarraj, A.; Annlin Jeba, S.V.; Praveena, V. Blockchain Technology in Smart-Cities. In *Blockchain Technology: Applications and Challenges*; Springer International Publishing: Cham, Switzerland, 2021; pp. 179–200. [CrossRef]
25. hacen Diallo, E.; Dieye, M.; Dib, O.; Valiorgue, P. An agnostic and secure interoperability protocol for seamless asset movement. *J. Netw. Comput. Appl.* **2024**, *230*, 103930. [CrossRef]
26. Bhattacharya, S.; Chengoden, R.; Srivastava, G.; Alazab, M.; Javed, A.R.; Victor, N.; Maddikunta, P.K.R.; Gadekallu, T.R. Incentive Mechanisms for Smart Grid: State of the Art, Challenges, Open Issues, Future Directions. *Big Data Cogn. Comput.* **2022**, *6*, 47. [CrossRef]
27. Kassen, M. Understanding decentralized civic engagement: Focus on peer-to-peer and blockchain-driven perspectives on e-participation. *Technol. Soc.* **2021**, *66*, 101650. [CrossRef]
28. Safety Reports. Safety-Reports Incident Reporting App. Available online: https://www.safety-reports.com/safety-incident-app/ (accessed on 29 February 2024).
29. Trackforce Valiant. Trackforce Valiant Homepage. Available online: https://www.trackforcevaliant.com/ (accessed on 29 February 2024).
30. Risk Assessor. Free Incident Reporting Software. Available online: https://www.riskassessor.net/incident-reporting (accessed on 29 February 2024).

31. Petschnig, W.; Haslinger-Baumann, E. Critical Incident Reporting System (CIRS): A fundamental component of risk management in health care systems to enhance patient safety. *Saf. Health* **2017**, *3*, 9. [CrossRef]
32. Sendlhofer, G.; Schweppe, P.; Sprincnik, U.; Gombotz, V.; Leitgeb, K.; Tiefenbacher, P.; Kamolz, L.P.; Brunner, G. Deployment of Critical Incident Reporting System (CIRS) in public Styrian hospitals: A five year perspective. *BMC Health Serv. Res.* **2019**, *19*, 412. [CrossRef]
33. Qureshi, A.; Megias, D.; Rifà-Pous, H. A survey on incident reporting and management systems. In Proceedings of the Reunión española Sobre Criptología y Seguridad de la Información (RECSI), Granada, Spain, 3–5 October 2018.
34. Nakamoto, S. Bitcoin: A Peer-to-Peer Electronic Cash System. 2008. Available online: https://bitcoin.org/en/ (accessed on 29 February 2024).
35. Antonopoulos, A.M. *Mastering Bitcoin: Programming the Open Blockchain*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2017.
36. Dib, O.; Brousmiche, K.L.; Durand, A.; Thea, E.; Hamida, E.B. Consortium blockchains: Overview, applications and challenges. *Int. J. Adv. Telecommun.* **2018**, *11*, 51–64.
37. Tapscott, D.; Tapscott, A. *Blockchain Revolution: How the Technology behind Bitcoin is Changing Money, Business, and the World*; Penguin: London, UK, 2016.
38. hacen Diallo, E.; Dib, O.; Al Agha, K. A scalable blockchain-based scheme for traffic-related data sharing in VANETs. *Blockchain Res. Appl.* **2022**, *3*, 100087. [CrossRef]
39. Marbouh, D.; Simsekler, M.C.E.; Salah, K.; Jayaraman, R.; Ellahham, S. Blockchain-based Incident Reporting System for Patient Safety and Quality in Healthcare. In *Trust Models for Next-Generation Blockchain Ecosystems*; Springer: Cham, Switzerland, 2021; pp. 167–190. [CrossRef]
40. Marbouh, D.; Simsekler, M.C.E.; Salah, K.; Jayaraman, R.; Ellahham, S. A Blockchain-Based Regulatory Framework for mHealth. *Data* **2022**, *7*, 177. [CrossRef]
41. Aktas, E.; Demir, S.; Paksoy, T. The use of blockchain in aviation safety reporting systems: A framework proposal. *Int. J. Aerosp. Psychol.* **2022**, *32*, 283–306. [CrossRef]
42. Kao, K.; Ahmed, S.; Pyala, R.; Alsabri, M. Incident Reporting Systems: How did we get here and where should we go? A narrative review. *Front. Emerg. Med.* **2022**, *6*. [CrossRef]
43. Loren, D.J.; Garbutt, J.; Dunagan, W.C.; Bommarito, K.M.; Ebers, A.G.; Levinson, W.; Waterman, A.D.; Fraser, V.J.; Summy, E.A.; Gallagher, T.H. Risk managers, physicians, and disclosure of harmful medical errors. *Jt. Comm. J. Qual. Patient Saf.* **2010**, *36*, 101–108. [CrossRef]
44. Jha, P. If no-one helps you after a car crash in India, this is why. *BBC News*, 7 June 2016.
45. Macrae, C. The problem with incident reporting. *BMJ quality & safety* **2016**, *25*, 71–75. [CrossRef]
46. Donaldson, L. When will health care pass the orange-wire test? *Lancet* **2004**, *364*, 1567–1568. [CrossRef]
47. Vogus, T.J. *Close Calls: Managing Risk and Resilience in Airline Flight Safety*; Springer: Berlin/Heidelberg, Germany, 2018. [CrossRef]
48. Kingston, M.J.; Evans, S.M.; Smith, B.J.; Berry, J.G. Attitudes of doctors and nurses towards incident reporting: A qualitative analysis. *Med. J. Aust.* **2004**, *181*, 36–39. [CrossRef]
49. Evans, S.M.; Berry, J.G.; Smith, B.J.; Esterman, A.; Selim, P.; O'Shaughnessy, J.; DeWit, M. Attitudes and barriers to incident reporting: A collaborative hospital study. *BMJ Qual. Saf.* **2006**, *15*, 39–43. [CrossRef]
50. Javed, A.R.; Shahzad, F.; ur Rehman, S.; Zikria, Y.B.; Razzak, I.; Jalil, Z.; Xu, G. Future smart cities: Requirements, emerging technologies, applications, challenges, and future aspects. *Cities* **2022**, *129*, 103794. [CrossRef]
51. Jilani, M.T.; Ur Rehman, M.Z.; Abbas, M.A. An Application Framework of Crowdsourcing based Emergency Events Reporting in Smart Cities. In Proceedings of the 2019 International Symposium on Networks, Computers and Communications (ISNCC), Istanbul, Turkey, 18–20 June 2019; pp. 1–5. [CrossRef]
52. Suciu, G.; Necula, L.A.; Jelea, V.; Cristea, D.S.; Rusu, C.C.; Mistodie, L.R.; Ivanov, M.P. Smart City Platform Based on Citizen Reporting Services. In *Advances in Industrial Internet of Things, Engineering and Management*; Springer International Publishing: Cham, Switzerland, 2021; pp. 87–100. [CrossRef]
53. Zheng, G.; Gao, L.; Huang, L.; Guan, J. Decentralized Application (DApp). In *Ethereum Smart Contract Development in Solidity*; Springer: Singapore, 2021; pp. 253–280. [CrossRef]
54. Dib, O.; Toumi, K. Decentralized identity systems: Architecture, challenges, solutions and future directions. *Ann. Emerg. Technol. Comput. (AETiC)* **2020**, *4*, 19–40. [CrossRef]
55. Butincu, C.N.; Alexandrescu, A. Design Aspects of Decentralized Identifiers and Self-Sovereign Identity Systems. *IEEE Access* **2024**, *12*, 60928–60942. [CrossRef]
56. Shcherbakov, A. Hyperledger Indy Besu as a permissioned ledger in Selfsovereign Identity. In *Open Identity Summit 2024*; Gesellschaft für Informatik eV: Bonn, Germany, 2024; pp. 127–137.
57. Dieye, M.; Valiorgue, P.; Gelas, J.P.; Diallo, E.H.; Ghodous, P.; Biennier, F.; Peyrol, E. A Self-Sovereign Identity Based on Zero-Knowledge Proof and Blockchain. *IEEE Access* **2023**, *11*, 49445–49455. [CrossRef]
58. Naik, N.; Jenkins, P. Self-Sovereign Identity Specifications: Govern Your Identity Through Your Digital Wallet using Blockchain Technology. In Proceedings of the 2020 8th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud), Oxford, UK, 3–6 August 2020; pp. 90–95. [CrossRef]
59. Viriyasitavat, W.; Xu, L.D.; Niyato, D.; Bi, Z.; Hoonsopon, D. Applications of Blockchain in Business Processes: A Comprehensive Review. *IEEE Access* **2022**, *10*, 118900–118925. [CrossRef]

60. Li, Y.; Fan, Y.; Zhang, L.; Crowcroft, J. RAFT Consensus Reliability in Wireless Networks: Probabilistic Analysis. *IEEE Internet Things J.* **2023**, *10*, 12839–12853. [CrossRef]
61. Zou, W.; Lo, D.; Kochhar, P.S.; Le, X.B.D.; Xia, X.; Feng, Y.; Chen, Z.; Xu, B. Smart Contract Development: Challenges and Opportunities. *IEEE Trans. Softw. Eng.* **2021**, *47*, 2084–2106. [CrossRef]
62. Benet, J. *IPFS—Content Addressed Versioned P2P File System*; Technical report; Protocol Labs Inc.: San Francisco, CA, USA, 2014.
63. Dib, O.; Nan, Z.; Liu, J. Machine learning-based ransomware classification of Bitcoin transactions. *J. King Saud Univ.-Comput. Inf. Sci.* **2024**, *36*, 101925. [CrossRef]
64. Liu, J.; Xue, H.; Wang, J.; Hong, S.; Fu, H.; Dib, O. A Systematic Comparison on Prevailing Intrusion Detection Models. In Proceedings of the International Conference on Parallel and Distributed Computing: Applications and Technologies, Sendai, Japan, 7–9 December 2022; Springer: Cham, Switzerland, 2023; pp. 213–224. [CrossRef]
65. ConsenSys. Quorum: A Permissioned Implementation of Ethereum Supporting Data. 2016. Privacy. Available online: https://github.com/ConsenSys/quorum/blob/master/docs/Quorum%20Whitepaper%20v0.2.pdf (accessed on 16 May 2024).
66. Abdullah Lajam, O.; Ahmed Helmy, T. Performance Evaluation of IPFS in Private Networks. In Proceedings of the 2021 4th International Conference on Data Storage and Data Engineering, Barcelona, Spain, 18–20 February 2021; pp. 77–84. [CrossRef]
67. Marion, J. Les Statistiques des Services d'Incendie et de Secours. 2023. Available online: https://www.interieur.gouv.fr/Publications/Statistiques/Securite-civile/2022 (accessed on 16 May 2024).
68. Mazzoni, M.; Corradi, A.; Di Nicola, V. Performance evaluation of permissioned blockchains for financial applications: The ConsenSys Quorum case study. *Blockchain Res. Appl.* **2022**, *3*, 100026. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

urban science

MDPI

Article

Air Quality Awareness—Empirical Evidence from a Comparative Perspective between Two European Cities

Ornella Salimbene [1,*], María Teresa Baeza-Romero [2], Francesco Pilla [3] and Gregor Čok [1]

[1] Chair of Spatial Planning, Faculty of Civil and Geodetic Engineering, University of Ljubliana, Jamova Cesta 2, 1000 Ljubljana, Slovenia; gregor.cok@fgg.uni-lj.si
[2] Department of Physical Chemistry, School of Industrial and Aerospace Engineering, Institute of Nanoscience, Nanotechnology and Molecular Materials, Universidad de Castilla-La Mancha, 45071 Toledo, Spain; mariateresa.baeza@uclm.es
[3] School of Architecture, Planning and Environmental Policy, UCD, D04 V1W8 Dublin, Ireland; francesco.pilla@ucd.ie
* Correspondence: ornella.salimbene@fgg.uni-lj.si

Abstract: This research, conducted as part of the European project PA-MAP (Participatory Approach to Monitoring Air Quality in Urban Environments), investigates citizens' perceptions of air pollution in two European cities: Turin (Italy) and Dublin (Ireland). We systematically collected social data through semi-structured surveys distributed via online platforms. Our aim was to gather comprehensive information on the perception of pollution sources and the effectiveness of control systems. We examined citizens' habits and their suggestions to improve air quality. The results revealed the following: (1) In Turin, the perception of air pollution significantly influences individuals' self-assessments of health. Consequently, increased vigilance is required to mitigate environmental health risks and to disseminate information on air quality and effective practices for reducing air pollution. (2) The use of sustainable transportation in Turin is hindered due to the inadequate maintenance of cycling infrastructure. Citizens prefer to use cars or trains, which facilitate rapid travel. (3) In Dublin, greater sensitivity to the problem of air pollution has emerged, and the use of sustainable means of transport is facilitated via ad hoc infrastructure. These results provide empirical evidence to policymakers to support the promotion of concrete local actions through which to improve air quality.

Keywords: air quality; citizen behavior; qualitative survey; Europe

1. Context and Motivation

Air pollution constitutes a major environmental problem and presents continuous risks to human health. Scientific evidence links air pollution to a range of acute and chronic diseases, including cardiovascular diseases, stroke, respiratory diseases, and cancer (World Health Statistics Report 2021) [1]. The European Environment Agency (Air Quality Report, EEA 2022) reported 238,000 premature deaths in Europe in 2020, while the World Health Organization estimates approximately 6.7 million premature deaths worldwide [2]. Extensive research consistently identifies air pollution as the fourth most significant health risk factor [3–5], and current medical evidence highlights that both short-term and long-term exposure to air pollution can lead to acute and chronic health conditions [6–8].

The Air Quality Index (AQI) serves as a vital parameter for assessing the impact of air quality on individual well-being [9,10]. However, in an era of increasing urbanization, citizens' perceptions can be a crucial indicator for evaluating the effectiveness of environmental policies and planning targeted interventions in social, environmental, and clinical sectors. In recent years, with the rise in research on risk perception, air pollution, as an environmental risk, has garnered significant attention from scientists [11,12].

Consequently, several European projects have been funded to improve air quality in Europe [13–15]. Building sustainable cities requires a strong commitment to addressing

environmental problems that adversely affect citizens' health. National protocols and the European Green Deal include a series of measures aimed at reducing greenhouse gases, improving energy efficiency, and protecting biodiversity. Additionally, the ambition to achieve climate neutrality by 2050 necessitates a progress-monitoring system that ensures a socio-economically efficient green transition and a consequent reduction in air pollution to mitigate its impact on public health. Empirical research on the perception of air pollution is needed, and it should involve specialists from different sectors, expanding investigations and data collection to the clinical, urban, and social spheres.

Addressing air pollution shares similarities with tackling other adverse effects in urban environments, such as heat islands, noise, and light pollution. The complexity of this issue necessitates a comprehensive understanding, as implementing effective solutions relies on both the availability of reliable data and strategic decision-making in urban development. Preventive and remedial measures can be designed at various scales—individual buildings, neighborhoods, and entire cities. However, these decisions require significant time and resources. Therefore, any research that provides policymakers with credible evidence on air pollution, its sources, and potential solutions is crucial.

Despite the priority given to these issues, interdisciplinary studies that examine how social, clinical, and environmental data influence citizens' perceptions of air quality remain scarce. The existing literature on this topic explores citizens' perceptions across various regions, contexts, and methodologies. For instance, a study conducted in China investigated the perception of air pollution and its impact on residents' health. Findings revealed that perceptions of air pollution significantly affect both self-rated health and mental health, underscoring the need to consider public perception as a critical component in managing air pollution and developing health policies [16]. Additionally, a survey conducted in Portugal assessed perceptions of air quality among the general population and a specific subgroup from an industrial area in the Lisbon metropolitan area. This study found that 61% of residents in the industrial area rated the air quality as "poor", compared to only 14% of the general population, highlighting heightened concern about air pollution in industrial areas relative to less polluted ones [17]. Current research predominantly focuses on developed countries [18–22], with a notable gap in studies from developing countries, where challenges include limited resources and technology. Some researchers have suggested using direct interviews in various European capitals [23–25]. For example, a study in Brussels collected social data through direct offline interviews, but this approach has several limitations, including a restricted sample size, high costs, time consumption, logistical complexity, and the constraints of conducting interviews in limited areas.

The literature indicates a need for social analyses that incorporate strategic indicators and the georeferencing of participants, which would allow for correlating citizens' perceptions with objective local concentration data. Additionally, the review highlights that studies on the effects of air pollution on physical and mental health often occur independently, leading to incomplete and non-systematic research. Contextual variations, such as differing air quality standards, further limit the explanatory power of findings in developing countries. Regarding social analyses of air quality perception, scientific studies have primarily focused on developed regions [26–28].

Assessing the impact of air pollution on residents' health through both offline and online social questionnaires, as well as co-creative actions, is of practical significance. Such assessments facilitate the evaluation of health risks associated with air pollution, promote healthy living environments and sustainable practices, enhance urban mobility, and aid in the prevention of related pathologies [29,30].

Main Findings and Contributions

Despite these efforts, air quality issues persist in many European cities [31,32], underscoring the necessity for new indicators based on social analysis and citizens' perceptions to improve environmental policies and urban living conditions. Specifically, to our knowledge, no social surveys have been conducted in the cities examined in this study that include

specific questions about control systems, the impact of pollutants (such as particulate matter and black carbon), and personal actions taken to improve air quality. This research, conducted within the PA-MAP project, addresses these gaps by presenting findings from quantitative and qualitative investigations into urban air quality perception in Turin (Italy) and Dublin (Ireland).

Since the implementation of recent policy reforms, Italy and Ireland have experienced rapid economic development, leading to persistent environmental pollution issues, particularly in Northern Italy [33,34]. The health risks associated with air pollution have become a major public concern in these regions, prompting several institutional agreements aimed at improving air quality [35]. The proposed survey serves as a crucial resource for providing policymakers with detailed information on air quality and traffic management in both an industrial-Mediterranean context and a Nordic urban context.

This research seeks to augment the limited body of qualitative research in this field by examining perception from various complementary perspectives. It is divided into four sections: demographic information, air quality perception, trust in institutions, and suggestions to improve air quality, including the use of cycling infrastructure. The study explores how cultural, political, and demographic factors influence citizens' opinions and proposes tailored strategies to address environmental challenges specific to each city. One innovative aspect of this research is the collection of concrete suggestions related to air-monitoring systems and cycling infrastructure. The data gathered provide unique insights into key factors that either hinder or promote cycling.

This comparative and participatory approach not only evaluates citizens' perceptions but also offers practical recommendations based on empirical data to improve air quality and sustainable mobility in the cities studied. The research also aims to investigate citizens' trust in local institutions. It employs qualitative methods, chi-square tests, and basic statistical analyses to assess how age and profession influence air quality perceptions, information levels, and communication channels used by citizens.

Additionally, the research examines two critical aspects: first, the actions citizens are willing to take to improve air quality; and second, proposals for enhancing air quality within their cities. Citizens' opinions on air quality were compared with objective local air quality data published by environmental protection authorities. The study also considered local infrastructure that supports the use of sustainable transport, with urban cycling infrastructure analyzed using OpenStreetMap data [36]. The findings indicate that inadequate cycling infrastructure discourages bicycle use, and a heightened perception of air pollution correlates with worsened self-assessed health and increased health concerns. Citizens tend to trust institutions more when there is a higher presence of monitoring stations. These results have significant implications for research and policy aimed at improving air quality and public health, emphasizing the need to address non-users' motivations for sustainable transport.

This social investigation is highly relevant to both the academic community and policymakers. The findings highlight several key aspects:

1. Informing policy decisions: The collection of social data enables informed decision-making and the development of policies that address citizens' concerns. For example, our research identifies issues related to cycling infrastructure in Turin. Negative perceptions of infrastructure, urban spaces, or air quality should prompt the city's administration to implement more rigorous measures that promote and encourage the use of bike-sharing systems or private bicycles.
2. Evaluating policy effectiveness: The collected data serve as a foundation for assessing the effectiveness of existing policies. For instance, the positive perception of air quality and the quality of the cycling infrastructure network among Irish citizens indicate that the current policy measures are functioning well and should be continued and reinforced.
3. Identifying research priorities: The results suggest areas for further research. This study highlights specific concerns in the two cities that could lead to new investiga-

tions into the causes and effects of these perceptions. Future studies could adopt more comprehensive and inclusive methodological approaches, such as online and offline questionnaires and the use of low-cost sensors in various city areas.

4. Community involvement: The survey positively impacts community engagement by involving citizens in the decision-making process. When individuals feel heard, they are more likely to support and adhere to the policies implemented.

Overall, the proposed study provides an empirical foundation for improving public policies, assessing their impacts, and guiding future research, thereby contributing to the enhancement of quality of life in the studied urban contexts. These results should encourage local policymakers to intensify efforts to combat air pollution and protect residents' health. The comparison between two European cities, Turin and Dublin, offers a unique perspective on how different infrastructures and local sensitivities influence citizens' perceptions and the adoption of sustainable transportation practices. The research specifically highlights how the perception of pollution in Turin affects self-assessed health, an area of study that is not widely explored. The analysis of infrastructural barriers in Turin and infrastructural facilitations in Dublin provides practical and applicable insights for policymakers, based on empirical evidence gathered through a participatory approach. The research specifically highlights how perceptions of pollution in Turin influence self-rated health, an area of study that is not widely explored. It highlights the psychological and subjective dimensions of environmental health, which are often overlooked in traditional air quality studies.

The analysis of infrastructure barriers in Turin and infrastructure facilitators in Dublin provides practical and applicable insights for policymakers, based on empirical evidence gathered through a participatory approach. This evidence-based approach not only increases the credibility of the findings but also ensures that recommendations are based on real-world experiences and citizen feedback. This study stands out for its comparative approach, which reveals how local contexts shape environmental perceptions and behaviors. By integrating health perceptions and infrastructure analysis, it provides a comprehensive view that can guide targeted policy interventions. The participatory methodology further enriches the research, ensuring that the findings are not only theoretical but also practically relevant and actionable.

The research is structured in the following order: context and motivation, materials and methods, results and discussion, and conclusions and future work.

2. Materials and Methods

A transnational comparison of air quality perception serves as a valuable tool for observing differences and similarities across various geographical, cultural, and regulatory contexts.

A qualitative analysis based on a semi-structured, anonymous questionnaire was conducted to investigate the level of public awareness regarding air pollution in two European cities: Turin and Dublin. Turin was selected as a study area for two primary reasons: (1) its unique meteorological and geographic conditions in the Po Valley, which is one of the most polluted regions in Europe during both summer and winter [37], and (2) its high density of vehicular emissions, which ranks among the highest in Europe [38]. In contrast, Dublin is one of the capitals with high levels of citizen satisfaction regarding both quality of life and the quality of public transport [39–41].

The data collection process involved using an online form, and the interview protocols adhered to the legal regulations on the protection of personal data specified in Legislative Decree 196/2003 [42]. Additionally, these protocols were updated according to the new Legislative Decree 101/2018 [43] to align Italian legislation with the European privacy regulation [44]. Before its implementation, the survey underwent preliminary testing by the research team to validate its reliability and ensure clarity and understandability. Closed and open questions were used to gather suggestions from citizens. The questionnaires were disseminated via email and on the websites of local cultural institutes, remaining available online for six months during 2023. To minimize nonresponse bias, various strategies were

employed, including creating simple and clear questions and maintaining a survey length that allowed responses to be completed within a maximum of eight minutes. Data collection targeted a stratified population of residents in Turin and Dublin aged over 18 years. As previously mentioned, the questionnaire was administered via email, which naturally excluded segments of the population who were not using the internet. This limitation could be addressed in future studies by incorporating both online and offline data collection methods. The survey comprised 20 questions divided into four main sections:

- Section 1: demographic information (gender, age, education, and profession).
- Section 2: perception of air quality (using the Likert scale model for various aspects and including open questions for further comments).
- Section 3: suggestions to improve air quality (open questions).
- Section 4: the use of cycling infrastructure (closed and open questions).

The investigations also aimed to identify the primary sources of air pollution in the city and determine the main communication channels residents use to access real-time updates on local air quality. Statistical and thematic analyses enabled the comparison of citizens' perceptions and the collection of concrete suggestions to improve the environmental situation and promote the use of sustainable transportation in the two cities.

Citizens were asked about their health-related concerns and their level of satisfaction with local political management regarding urban air pollution prevention. The pilot study facilitated the investigation of citizens' trust in local institutions and the primary means of transport used for short and long journeys. Moreover, the study examined disparities between residents' perceptions of air pollution sources and the officially recognized primary causes of air pollution reported by government authorities.

3. Results and Discussion

The results obtained are summarized in Tables 1 and 2. Table 1 provides an overview of the demographic data collected from participants in the two European cities. Table 2 presents both crude odds ratios derived from bivariate logistic regression analyses and adjusted odds ratios from the multivariable logistic regression model, along with 95% confidence intervals and *p*-values.

Table 1. Demographics data, absolute frequency, and percentage values.

Place of Residence	Dublin City & Countryside (%)	Frequency	Torino City & Countryside (%)	Frequency
Gender				
Male	42.4	163	44.8	172
Female	57.2	220	54.9	211
Age				
19–28	29.9	115	23.9	92
29–39	34.4	132	32.03	123
40–65	27.3	105	39.06	150
>65	7.8	32	4.9	19
Education level				
Elementary	/		0.78	3
Secondary school	23.9	92	8.07	31
Hight school	30.7	118	23.9	92
University degree	41.9	161	59.1	227
Other	3.4	13	7.8	31
Job				
Student	20.3	78	15.3	59
Self-Employed	31.5	121	15.1	58
Employee	39.3	151	54.4	209
Unemployed	8.6	33	5.7	22
Other	0.26	1	9.1	36

Table 2. Age and educational level. Demographics data, absolute frequency, and percentage values.

	Crude OR			Adjusted OR		
	OR	**95%CI**	***p*-Value**	**OR**	**95%CI**	***p*-Value**
Age						
Age 40 to Age <40	0.48	0.48, 0.80	0.006	0.34	0.34, 0.9	0.035
Educational level						
High School (HS) to <HS	1.63	1.63, 3.5	0.00002	1.15	1.15, 3.02	0.015
College to <HS	2.70	2.70, 5.1	<0.00001	1.5	1.5, 3.81	0.0003

As shown in Table 1, the study involved citizens aged between 19 and 65 from Turin and Dublin. For comparative purposes, a total of 768 questionnaires were collected, with 384 responses from each city. Notably, 100% of the questionnaires met the study criteria, with each participant providing one response and completing at least 85% of the questions. In Turin, 39.1% of the responses were from citizens aged between 40 and 65, while in Dublin, participation by age group and education level was more evenly distributed. Among the respondents, Irish citizens aged between 29 and 39 accounted for 34.4% of the responses. Additionally, Irish students were highly engaged, contributing 20.3% of the responses, compared to 15.3% in the Turin sample. Female participants showed higher engagement, making up 57.2% of the respondents in Dublin and 54.9% in Turin. Overall, the pilot study had a greater representation of women (432) compared to men (336). Notably, young Irish individuals aged between 19 and 28 demonstrated greater awareness of air pollution issues (29.9%) compared to their Turin counterparts (23.9%) (see Table 2). This difference could be attributed to more robust air quality initiatives within the Irish university context. Furthermore, Dublin recorded a higher participation of older individuals (7.8%) compared to Turin (4.7%). Specifically, Dublin supplied a total of 247 participants under the age of 40, while Turin supplied 215 participants in the same age bracket. All participants reported residing in urban areas. Among the respondents in Dublin and Turin, the majority held a university degree (41.9% vs. 59.1%) or a high school diploma (30.7% vs. 23.9%). The respondents' employment status varied: 39.3% of Dublin respondents were employed, while 31.5% were self-employed. In contrast, Turin showed a significant difference, with only 15.1% working in the private sector and 54.4% being state employees. The percentage of unemployed participants was higher in Dublin (8.5%) compared to Turin (5.7%). Chi-square tests of independence were used to explore associations between participants' age, gender, education level, and beliefs regarding sources of air pollution and preferred information channels. A *p*-value of less than 5% was considered statistically significant. The bivariate analyses depicted in Table 2 showed that participants' age and education level were associated with air pollution awareness at a 5% significance level. The survey results highlight a robust and non-random relationship between these factors and the perception of air pollution.

3.1. Citizen's Perception of Air Pollution

In Turin, 42% of interviewees reported that the air quality was poor, while 36% perceived it as dangerous. Only 19% considered the air to be healthy. Conversely, in the Irish capital, the situation was markedly different. Approximately 72.6% of the Irish respondents asserted that they breathed good air, while 22.2% believed they inhaled polluted air. Only 3% of respondents described Dublin's air quality as mediocre. Figure 1 graphically summarizes the findings regarding air quality perception in these two European cities.

Among the respondents in Turin, there was notable discontent and mistrust towards institutions. This sentiment may partly explain the emergence of several associations and independent movements in Turin in recent years, which aim to raise awareness about air pollution and monitor specific gaseous pollutants using passive sensors distributed among citizens [45]. One prominent organization is "Torino Respira", a volunteer initiative focused on improving air quality through public events and the collection of environmental data via passive sensors. It is important to highlight that this study intentionally excluded environmental activists and members of private and public associations from the qualitative

investigation to avoid influencing the results. During the survey, the participants were asked to identify the primary sources of pollution in their cities from a list of options, including construction, incineration, heating, energy production, and transport. In Turin, the respondents predominantly identified three main contributors to air pollution: transport (58.3%), heating systems (18.2%), and energy emissions from production plants (11.45%). In contrast, the Dublin respondents primarily identified transport as the main source of pollution (66.6%), followed by heating (16.4%). Incineration plants (6.25%) and energy production facilities (5.6%) were reported as lesser contributors (see Figure 2).

(**a**) (**b**)

Figure 1. Participants' perception of air quality in Turin and Dublin.

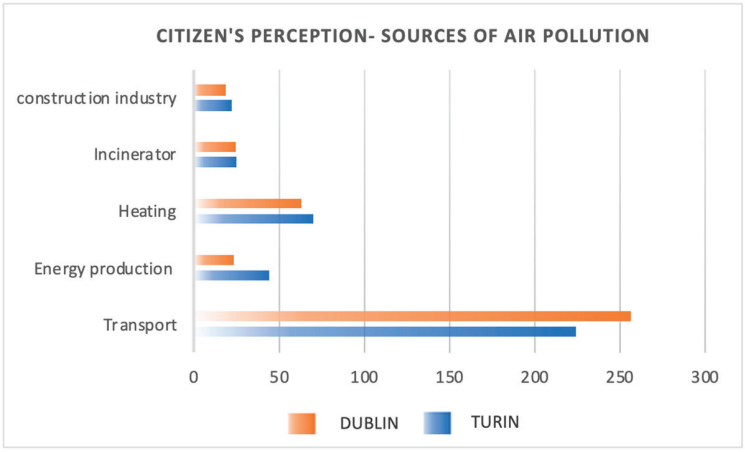


Figure 2. Citizens' perception of the main sources of air pollution in Dublin and Turin.

According to data published by the Municipality of Turin, 87% of the primary PM10 emissions are associated with transportation, with 38% of total emissions coming from transportation and 49% from heating systems [46]. Additionally, regional and municipal authorities periodically release reports on major sources of pollution, which are available online and include data on sources, concentration trends, and details about monitoring campaigns [47]. In Dublin, some studies suggest that traffic emissions have a less significant impact on residential areas compared to heating emissions. Various articles have addressed the effects of traffic on air quality and the variability of NOx levels in road canyons [48–50]. Data on air pollution sources in the Irish capital are also accessible through the EPA and municipal websites, which outline Ireland's national environmental policies and integrated policy frameworks [51].

Our survey results align with the existing literature, showing that transportation is the predominant contributor to urban air pollution, with heating systems also playing a substantial role. Participants in our study exhibited a strong awareness of these major pollution sources, consistent with the findings of local authorities. Despite this awareness, overcoming entrenched behaviors and encouraging the adoption of sustainable transport options in Turin remain a significant challenge. These challenges are compounded by issues related to cycling infrastructure, which we explore further below.

3.2. Means of Transport Used and Commuting

The survey aimed to evaluate citizens' preferences for sustainable and low-pollution modes of transportation during daily commutes (see Table 3). In Turin, 58.1% of the respondents reported using bicycles or walking exclusively for short trips. In contrast, Dublin showed a higher percentage, with 67.5% favoring these eco-friendly options. Despite the significant proportion of bicycle users in Turin, 54.9% indicated that they used bicycles solely for short journeys (less than 1 h), unlike in Dublin. In the Irish capital, citizens demonstrated a greater commitment to sustainable transportation. Bicycles were identified as the primary mode of commuting to work and school, with 30.8% of respondents using them for approximately 3 h daily, compared to around 10% in Turin.

Table 3. Means of transport used by the interviewees in Turin and Dublin during their daily commutes, as well as the time of use and motivation for their choice of means of transport.

		TORINO (%)	DUBLIN (%)
	Do you travel by bike or on foot in your city?		
	Yes	58.1	67.5
	Not	19.3	14.5
	Rarely	22.6	17.9
	If yes, for how many hours a day?		
	1 h	54.9	34.2
	2 h	28.4	20.5
	3 h	10	30.8
	4 h	4.1	12
	>4 h	1.9	2.6
	What means of transport do you use most regularly?		
	Car	35	11.1
	Bike	17.3	41.9
	Bus	15.6	29.1
	Subway	10.9	2.9
	Train	2.4	7.7
	I walk	16	7
/	Other	2.7	1.7
	Why do you use this means of transport?		
	It is comfortable	45.5	33.3
	It is cheap	7.9	12.8
	It is fast	19.9	17.9
	Doesn't pollute	15.1	27.4
/	Other	11.6	8.5

Additionally, 41.9% of respondents in Dublin primarily used bicycles for commuting to work, whereas only 17.3% of those in Turin preferred bicycles for this purpose. In Turin, car usage for work-related commutes was predominant, with 35% of respondents indicating it as their primary mode of transport. This was followed by bicycles (17.3%), walking (16%), buses (15.6%), the subway (10.9%), and other options, such as electric scooters (2.7%). In Dublin, cars ranked third at 11.1%, behind bicycles and buses, which accounted for 29.1%. When asked to explain their choice of the most frequently used mode of transport, respondents from both cities cited comfort and practicality as the key factors influencing their decisions.

Dublin's well-established network of cycle paths significantly enhances citizens' mobility, while Turin faces ongoing challenges due to an underdeveloped cycling infrastructure. Irish respondents displayed a strong environmental consciousness, highlighting the importance of using low-impact transportation methods. Specifically, 27.4% of Dublin participants expressed a strong aversion to pollution and demonstrated greater environmental sensitivity compared to their Turin counterparts. Conversely, respondents in Turin prioritized comfort and speed when selecting transportation options, with sustainability considerations being less prominent, ranking third at 15.1%. Table 3 provides a summary of the key findings from the interviews conducted in Turin and Dublin regarding transportation preferences.

3.3. Attitudes Toward Governmental Policy for the Management of Urban Air Quality and Citizens' Suggestions

During the survey, all participants were asked about their awareness of air quality-monitoring systems in their respective cities. In Turin, 31% of respondents reported being uninformed about air quality monitoring, whereas only 12.8% of Dublin respondents indicated a lack of awareness. Both cities identified websites and radio as the primary sources of information. However, in Dublin, 21.4% of respondents reported using specialized apps for air quality information. Regarding fixed monitoring stations, over half of the participants were aware of such stations' existence. In Dublin, 33.9% of respondents believed these stations provided an adequate analysis of the city's air quality. In contrast, Turin displayed significant skepticism and dissatisfaction with institutional systems. Specifically, 46% of Turin interviewees expressed dissatisfaction with the effectiveness of these systems, and 62% felt that PM concentration limits were not being adhered to. These findings may be influenced by negative media coverage of air quality issues in the Po Valley. It is difficult to determine whether Turin residents' perceptions of risk are due to high pollutant levels or influenced by media reports. In Turin, 68.7% of participants expressed health concerns related to air pollution and acknowledged inhaling unhealthy air, a situation markedly different from Dublin. Although Irish respondents were aware of the health effects of air pollution, only 35% expressed concern. Table 4 summarizes these findings.

Table 4. Participants' knowledge of air quality-monitoring systems and their efficiency, as well as the perception of health risk and compliance with limits.

	TURIN (%)	DUBLIN (%)
Do you check the air quality information in you city?		
Not	31	12.8
Web	27.6	31.6
Radio	23.9	31.6
App	9.1	21.4
I am an activist	3.4	0.9
Other	5.1	1.7
Do you know that there are fixed PM monitoring stations in your city?		
Yes	57.5	53.3
No	42.5	44.7

Table 4. *Cont.*

	TURIN (%)	DUBLIN (%)
Do you think these control systems are enough to monitor air quality?		
Yes	10.3	33.9
Not	46	9.6
I don't know	43.6	56.5
Do you think that the legal limit (PM) are respected in your city?		
Yes	10.2	47
Not	62	/
I don't know	27.8	53
Do you think air pollution harms your health?		
Yes	68.7	35
Not	3	6
I don't know	15.2	29.9
Not a lot	10.4	28.2
Only the weak (Children and elderly)	2.7	0.9

3.4. Citizens' Attitudes Toward Local Government

The qualitative analysis explored participants' attitudes toward government policies designed to improve air quality and their trust in institutions. In Turin, only 5.8% of the respondents expressed satisfaction with the political management of urban air quality. In contrast, 42.7% of the respondents in Dublin reported contentment with local policies and administrative management. Notably, 47.9% of Dublin participants chose not to provide a definitive response but expressed strong confidence in institutional effectiveness (see Figure 3).

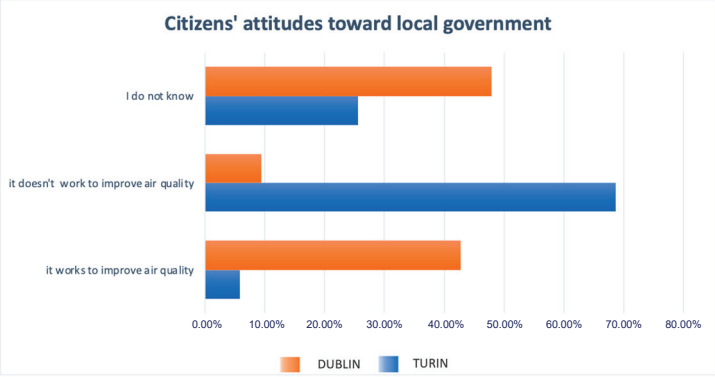


Figure 3. Participants' opinions regarding governmental action aimed at air pollution management.

The social survey incorporated questions aligned with the PA-MAP principles, specifically targeting "Actions for Change"—individual efforts aimed at mitigating and reducing air pollution. Figure 4 displays the responses from participants in Turin. Notably, 37% of respondents reported using energy-saving lamps, while 36% optimized their heating and cooling systems. Additionally, 9% preferred solid fuels, and 18% paid attention to sustainable modes of transport. In contrast, patterns in Dublin revealed different trends. In the Irish capital, 42% of respondents actively employed sustainable transportation methods, thereby lessening their impact on air quality. Furthermore, 22% prioritized optimizing heating systems, 20% were selective about their fuel choices, and 16% used energy-saving light bulbs. Remarkably, all interviewees provided insights into their personal actions.

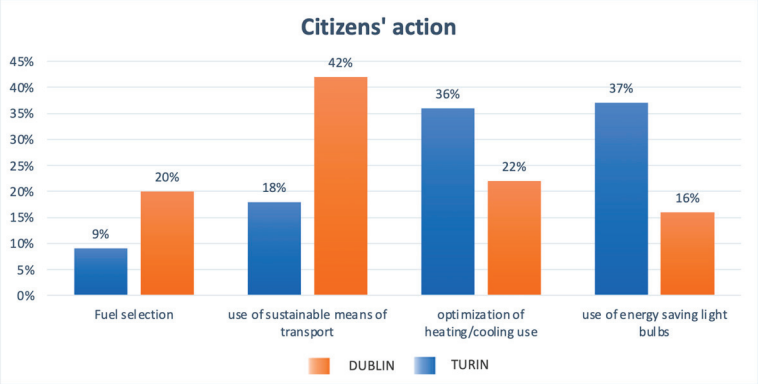


Figure 4. Citizens' actions to limit air pollution in their city.

During interviews conducted in Turin, 68.6% of respondents advocated for the implementation of additional preventive measures to enhance air quality in the city. Key recommendations included increasing the oversight of industrial facilities and emissions (28%) and urging improvements in urban infrastructure and cycling pathways (52%). Additionally, there were calls to restrict waste incineration (20%) (see Figure 5). In contrast, respondents in Dublin emphasized the need for the continued monitoring of wood and solid fuel usage. This focus aligns with the existing literature, which indicates that these factors contribute up to 70% of the ultrafine particulate matter concentrations recorded in the Irish capital [52]. Irish respondents showed a willingness to reduce their reliance on solid biomass and actively adopt sustainable transportation options, supported by effective infrastructure. According to interviewees in Dublin, transportation remained a significant concern, along with emissions from ports and industrial sources, which accounted for 37% of the expressed concerns.

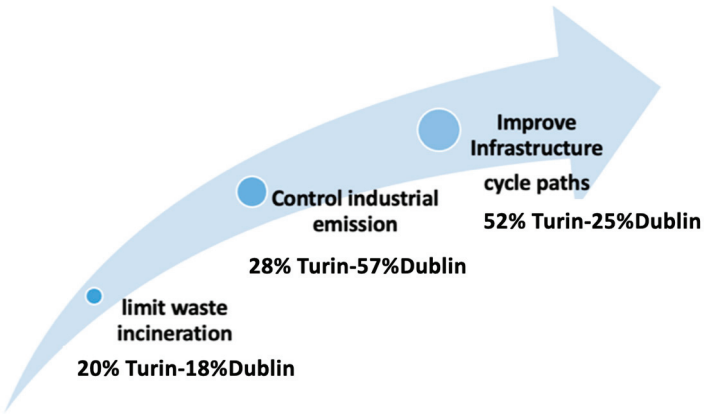


Figure 5. Participant suggestions: the top three areas requiring increased government attention according to citizens' recommendations.

The educational levels in both cities likely influenced the public's understanding of major sources of air pollution. Individuals with higher educational qualifications, such as degrees, demonstrated greater awareness and engagement. These findings are consistent with prior research, which shows that social and economic contexts can affect individuals' awareness of air pollution sources [53]. In Dublin, 27.4% of respondents reported using sustainable transportation to mitigate their contribution to air pollution. Participants from both Dublin and Turin identified urban industrial plants and waste incineration as significant sources of air pollution that require more effective government control. Studies conducted in Ireland further corroborate these public concerns, particularly regarding waste incineration [54].

Several studies have established a direct link between the proximity of incinerators to residential areas and increased mortality rates from certain cancers [55]. In this context, 20% of respondents in Turin expressed concerns about waste incineration. The study also explored participants' attitudes towards government actions related to environmental protection. A significant majority of interviewees in Turin (68.6%) felt that the government's efforts were insufficient, indicating a high level of institutional mistrust. Specifically, 52% of Turin participants suggested that the government should allocate more resources to infrastructure development and the enhancement of cycling lanes—an issue that also ranked as a high priority in Dublin. Conversely, 47.9% of Irish respondents expressed confidence in their local government, and 57% emphasized the need for stronger regulation of industrial emissions and better management of incineration facilities. Notably, while most interviewees in Dublin advocated for improved controls on industrial emissions and incineration, those in Turin focused more on securing funding for urban infrastructure improvements.

In Turin, approximately 201 km of cycle paths exist [56]. However, according to interviewees, the poor maintenance of these paths renders them less usable, which partly explains why many individuals avoid using bicycles for longer journeys. Despite extensive promotional campaigns advocating bicycle usage, the TO-bike bike-sharing stations in Turin were deactivated in February 2023. Nevertheless, other shared public transport systems remain operational. This situation contrasts sharply with Dublin, where City Council initiated the installation of cycle lanes across the city from the 1990s to 2012. Dublin now boasts over 200 km of on- and off-road cycle lanes, accommodating both short and long journeys. Notably, in 2011, Dublin secured the ninth position in the Copenhagenize index of bicycle-friendly cities worldwide. This favorable ranking likely contributes to the interviewees' satisfaction and their strong inclination to utilize bicycles. Dublin's self-service rental system, known as Dublin-bike, features 102 stations citywide with more than 1500 bikes, and it serves over 60,000 subscribers [57]. Since February 13, 2023, the TOBike bike-sharing service in Turin has been replaced with a range of alternatives, including electric scooters managed by Bird, Bolt, Dott, Helbiz, Lime, and Voi, as well as a free-floating bike-sharing service managed by Bolt, Lime, and RideMovi, which allows bicycles to be parked on the street in compliance with regulations. All information regarding sustainable mobility can be found online [58]. Despite the presence of various transport services in Turin, interviewees reported that these options are primarily suitable for short journeys.

4. Conclusions and Future Work

Air pollution has become a significant global challenge. This article presents empirical findings derived from social data on air pollution in two European cities, aiming to investigate citizens' perceptions of air quality. A chi-square analysis was performed to examine how age and education level correlate with awareness of air pollution, revealing how these factors can influence individuals' perceptions and understanding of this environmental issue. The research also assessed participants' knowledge of air quality-monitoring systems and their perceived effectiveness. The findings suggest a need for improved communication and increased awareness regarding the effectiveness of these monitoring systems. The survey was predominantly completed by women, with the highest levels of participation observed among individuals with higher educational qualifications (degrees). The data presented do not establish a direct causal relationship between gender and response behavior. To determine whether observed differences in response behavior are genuinely attributable to gender, rather than other influencing factors, it would be necessary to demonstrate that variables affecting response behavior are independent of gender. Additionally, the qualitative survey gathered information on participants' habits related to air quality and explored potential actions that could be recommended to local governments for enhancing urban air quality. The study revealed five key observations:

1. Monitoring network credibility: A functional and accessible monitoring network that provides real-time data is crucial for enhancing institutional credibility and reassuring citizens. The monitoring network in Turin should be made more pervasive,

accessible, and capable of delivering real-time information to keep the local population consistently informed about air quality.

2. Alignment of perceptions and objective data: There is a significant alignment between residents' perceptions of local air pollution and the objective data from the respective cities. In Turin, air pollution is a well-recognized social issue with notable health impacts on residents [59–61]. Future research should focus on exploring any discrepancies between perceived and objective data to identify and address critical local issues.
3. Health concerns and objective evidence: Citizens' concerns about health contribute to their perceptions of poor air quality. The literature provides objective data on the correlation between air pollution and respiratory diseases in Turin [62–65]. Understanding the relationship between citizens' health concerns and objective evidence of disease can help shape more effective public health policies.
4. Primary sources of air pollution: Survey participants identified transportation and domestic heating as the primary sources of air pollution in both Turin and Dublin. These findings are consistent with objectively measured pollution levels reported in the literature [66] and by local political authorities. Analyzing public perceptions of pollution sources can highlight areas for improvement in managing and communicating air quality.
5. Infrastructure and transport choices: The quality of infrastructure and cycling facilities significantly influences citizens' choices regarding alternative modes of transport for long distances. The study shows that 42% of respondents in Dublin use bicycles for commuting and long journeys. This underscores the importance of investing in dedicated cycling infrastructure to promote sustainable transport options such as bicycles and electric scooters. Improved cycling infrastructure not only supports healthier and more sustainable lifestyles but also delivers economic, environmental, and social benefits. These findings suggest that increased investment in cycling infrastructure could enhance daily transportation options in Turin.

The methodology and findings of this research provide a foundation for developing policies that encourage behaviors to improve overall air quality. Citizens' perspectives and perceptions regarding air pollution have significant potential to drive improvements at the neighborhood level. The study offers a substantial amount of data that can inform local policymakers. By leveraging qualitative investigations, researchers can move from individual viewpoints to a broader understanding, identifying gaps and strengths related to public use of urban spaces, infrastructure, and other factors. However, it is important to recognize a limitation of this study: the data collection relied on electronic services provided via universities and cultural institutes, which may have inadvertently excluded segments of the population from diverse social contexts. Future research should aim to create inclusive tools and criteria that engage a broader range of social groups experiencing various boundary conditions. In conclusion, the study yields several observations:

- Future research should incorporate real-time comparisons between actual pollution data and citizens' perceptions, potentially through a smartphone application that allows daily feedback on air quality from different city areas.
- The representativeness of the sample is crucial, as achieving a sample that accurately reflects the demographic and socioeconomic diversity of the local community remains challenging.
- Internet access limitations affect participation, as not all citizens have access to the digital technologies needed to complete online questionnaires, thereby excluding some population segments.
- Measures should be taken to balance different groups in data analysis in order to avoid unrepresentative or biased conclusions.

Addressing these critical issues and implementing appropriate solutions could lead to the development of a model for assessing citizens' satisfaction with air quality. Promoting awareness, enhancing environmental education, and aligning citizens' needs with their

perceptions of pollution sources can encourage local authorities to undertake more effective and concrete collaborative planning initiatives.

Author Contributions: Conceptualization, O.S.; methodology, O.S.; software, O.S.; validation, O.S., F.P., G.Č. and M.T.B.-R.; formal analysis, O.S.; investigation, O.S.; resources, O.S.; data curation, O.S. and M.T.B.-R.; writing—original draft preparation, O.S.; writing—review and editing, O.S., G.Č., M.T.B.-R. and F.P.; visualization, M.T.B.-R.; supervision, M.T.B.-R., F.P. and G.Č.; project administration, O.S.; funding acquisition, O.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Slovenian Research and Innovation Agency through the Research Programme "Geoinformation infrastructure and sustainable spatial development of Slovenia," No. P2-0227. This research was conducted as part of the PA-MAP Project, which has received funding from the European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie grant agreement No. 945380-EUTOPIA.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors upon request.

Acknowledgments: The authors thank the interviewees of the two cities involved in the survey.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. *World Health Statistics 2021: Monitoring Health for the SDGs, Sustainable Development Goals*; World Health Organization: Geneva, Switzerland, 2021.
2. WHO Global Air Quality Guidelines. *Particulate Matter (PM2.5 and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide*; World Health Organization: Geneve, Switzerland, 2021.
3. Duan, R.; Hao, K.; Yang, T. Air pollution and chronic obstructive pulmonary disease. *Chronic Dis. Transl. Med.* **2020**, *6*, 260–269. [CrossRef] [PubMed] [PubMed Central]
4. Cohen, A.J.; Brauer, M.; Burnett, R.; Anderson, H.R.; Frostad, J.; Estep, K.; Balakrishnan, K.; Brunekreef, B.; Dandona, L.; Dandona, R.; et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: An analysis of data from the Global Burden of Diseases Study 2015. *Lancet* **2017**, *389*, 1907–1918. [CrossRef] [PubMed]
5. Huang, J.; Pan, X.; Guo, X.; Li, G. Health impact of China's Air Pollution Prevention and Control Action Plan: An analysis of national air quality monitoring and mortality data. *Lancet Planet Health* **2018**, *2*, e313–e323. [CrossRef]
6. Hermayurisca, F.; Taneepanichskul, N. Estimation of premature death attributed to short-and long-term PM2.5 exposure in Thailand. *Environ. Monit. Assess.* **2023**, *195*, 1176. [CrossRef]
7. Gilardi, L.; Marconcini, M.; Metz-Marconcini, A.; Esch, T.; Erbertseder, T. Long-term exposure and health risk assessment from air pollution: Impact of regional scale mobility. *Int. J. Health Geogr.* **2023**, *22*, 11. [CrossRef]
8. Koch, S.; Hoffmann, C.; Caseiro, A.; Ledebur, M.; Menk, M.; von Schneidemesser, E. Air quality in Germany as a contributing factor to morbidity from COVID-19. *Environ. Res.* **2022**, *214*, 113896. [CrossRef]
9. Sokhi, R.S.; Moussiopoulos, N.; Baklanov, A.; Bartzis, J.; Coll, I.; Finardi, S.; Friedrich, R.; Geels, C.; Grönholm, T.; Halenka, T.; et al. Advances in air quality research—Current and emerging challenges. *Atmospheric Meas. Technol.* **2022**, *22*, 4615–4703. [CrossRef]
10. Priti, K.; Kumar, P. A critical evaluation of air quality index models (1960–2021). *Environ. Monit. Assess.* **2022**, *194*, 324. [CrossRef]
11. Wong, C.M.; Vichit-Vadakan, N.; Vajanapoom, N.; Ostro, B.; Thach, T.Q.; Chau, P.Y.; Chan, E.K.; Chung, R.Y.; Ou, C.Q.; Yang, L.; et al. Part 5. Public health and air pollution in Asia (PAPA): A combined analysis of four studies of air pollution and mortality. *Res. Rep. Health Eff. Inst.* **2010**, *154*, 377–418. [PubMed]
12. Yen, I.H.; Yelin, E.H.; Katz, P.; Eisner, M.D.; Blanc, P.D.; Yen, I.H.; Yelin, E.H.; Katz, P.; Eisner, M. Perceived neighborhood problems and quality of life, physical functioning, and depressive symptoms among adults with Asthma. *Am. J. Public Health* **2006**, *96*, 873–879. [CrossRef]
13. Improving the Smart Control of Air Pollution in Europe. Available online: https://www.iscapeproject.eu (accessed on 11 August 2024).
14. ESCAPE Publications. Available online: https://projectescape.eu/escape-publications (accessed on 11 August 2024).
15. Open-air Laboratories for Nature Based Solutions to Manage Hydro-Meteo Risks. Available online: https://www.operandum-project.eu/ (accessed on 11 August 2024).
16. Zhu, J.; Lu, C. Air Quality, Pollution Perception, and Residents' Health: Evidence from China. *Toxics* **2023**, *11*, 591. [CrossRef] [PubMed]
17. Canha, N.; Justino, A.R.; Gamelas, C.A.; Almeida, S.M. Citizens' Perception on Air Quality in Portugal—How Concern Motivates Awareness. *Int. J. Environ. Res. Public Health* **2022**, *19*, 12760. [CrossRef] [PubMed]
18. Bruce, N.; Perez-Padilla, R.; Albalak, R. Indoor air pollution in developing countries: A major environmental and public health challenge. *Bull. World Health Organ.* **2000**, *78*, 1078–1092. [PubMed] [PubMed Central]

19. Gordon, S.B.; Bruce, N.G.; Grigg, J.; Hibberd, P.L.; Kurmi, O.P.; Lam, K.B.; Mortimer, K.; Asante, K.P.; Balakrishnan, K.; Balmes, J.; et al. Respiratory risks from household air pollution in low- and middle-income countries. *Lancet Respir. Med.* **2014**, *2*, 823–860. [CrossRef] [PubMed] [PubMed Central]
20. Liu, H.; Hu, T. How does air quality affect residents life satisfaction? Evidence based on multiperiod follow-up survey data of 122 cities in China. *Environ. Sci. Pollut. Res.* **2021**, *28*, 61047–61060. [CrossRef]
21. Nali, C.; Lorenzini, G. Air quality survey carried out by schoolchildren: An innovative tool for urban planning. *Environ. Monit. Assess.* **2007**, *131*, 201–210. [CrossRef]
22. Pfleger, E.; Adrian, C.; Lutz, R.; Drexler, H. Science communication on the public health risks of air pollution: A computational scoping review from 1958 to 2022. *Arch. Public Health* **2023**, *81*, 14. [CrossRef]
23. Reames, T.G.; Bravo, M.A. People, place and pollution: Investigating relationships between air quality perceptions, health concerns, exposure, and individual-and area-level characteristics. *Environ. Int.* **2019**, *122*, 244–255. [CrossRef]
24. Noël, C.; Van Landschoot, L.; Vanroelen, C.; Gadeyne, S. The Public's Perceptions of Air Pollution: What's in a Name? *Environ. Health Insights* **2022**, *16*, 11786302221123563. [CrossRef]
25. Al-Shidi, H.K.; Ambusaidi, A.K.; Sulaiman, H. Public awareness, perceptions and attitudes on air pollution and its health effects in Muscat, Oman. *J. Air Waste Manag. Assoc.* **2021**, *71*, 1159–1174. [CrossRef]
26. Oltra, C.; Sala, R. Perception of risk from air pollution and reported behaviors: A cross-sectional survey study in four cities. *J. Risk Res.* **2018**, *21*, 869–884. [CrossRef]
27. Giannico, V.; Spano, G.; Elia, M.; D'este, M.; Sanesi, G.; Lafortezza, R. Green spaces, quality of life, and citizen perception in European cities. *Environ. Res.* **2021**, *196*, 110922. [CrossRef]
28. Maione, M.; Mocca, E.; Eisfeld, K.; Kazepov, Y.; Fuzzi, S. Public perception of air pollution sources across Europe. *Ambio* **2020**, *50*, 1150–1158. [CrossRef]
29. Gignac, F.; Righi, V.; Toran, R.; Errandonea, L.P.; Ortiz, R.; Nieuwenhuijsen, M.; Creus, J.; Basagaña, X.; Balestrini, M. Co-creating a local environmental epidemiology study: The case of citizen science for investigating air pollution and related health risks in Barcelona, Spain. *Environ. Health* **2022**, *21*, 11. [CrossRef]
30. Fu, J.; Fu, H.; Zhu, C.; Sun, Y.; Cao, H.; Deng, X. The impact of air pollution on residents health in China: The mediating effect of population migration. *Air Qual. Atmos. Health* **2024**, *17*, 1281–1293. [CrossRef]
31. Tankosic', J.V. Environmental policy and air quality standards of the European Union. *J. Agron. Technol. Eng. Manag.* **2023**, *5*, 818–825. [CrossRef]
32. Karavas, Z.; Karayannis, V.; Moustakas, K. Comparative study of air quality indices in the European Union towards adopting a common air quality index. *Energy Environ.* **2021**, *32*, 959–980. [CrossRef]
33. OECD. *OECD Environmental Performance Review: Ireland 2021. OECD Environmental Performance Reviews*; OECD Publishing: Paris, France, 2021. [CrossRef]
34. MATTM. *Environmental Challenges, Summary of the State of the Environment in Italy*; MATTM: Rome, Italy, 2009.
35. Padano Accord. Available online: https://www.salute.gov.it/imgs/C_17_notizie_1473_listaFile_itemName_0_file.pdf (accessed on 11 August 2024).
36. Open Street Map Data. Available online: https://www.openstreetmap.org/#map=6/40.007/-2.488 (accessed on 11 August 2024).
37. Deserti, M.; Raffaelli, K.; Ramponi, L.; Carbonara, C.; Agostini, C.; Amorati, R.; Arvani, B.; Giovannini, G.; Maccaferri, S.; Poluzzi, V.; et al. COVID-19 and Air Quality in the Po Valley. Life Prep-Air: Emilia-Romagna Region, Italy, 2020.
38. Padoan, E.; Ajmone-Marsan, F.; Querol, X.; Amato, F. An empirical model to predict road dust emissions based on pavement and traffic characteristics. *Environ. Pollut.* **2018**, *237*, 713–720. [CrossRef]
39. Sustainable Movement and Transport. Available online: https://www.nationaltransport.ie/wp-content/uploads/2023/02/2022-NTA-Customer-Satisfaction.pdf (accessed on 1 March 2023).
40. Available online: https://www.dublincity.ie/sites/default/files/2022-12/Final%201-08%20Sus%20Mov.pdf (accessed on 6 November 2022).
41. Citizen Survey Report Dublin's Night-Time Economy. Available online: https://www.dublincity.ie/sites/default/files/2023-09/Executive%20Summary%20Your%20Dublin%20Your%20Voice%20Night%20Time%20Economy%20Survey%20Q3%202023.pdf (accessed on 6 November 2022).
42. Legislative Decree 196/2003. Available online: http://www.privacy.it/archivio/privacycode-en.html (accessed on 11 August 2024).
43. Legislative Decree 101/2018. Available online: https://www.gazzettaufficiale.it/eli/id/2018/09/04/18G00129/sg (accessed on 11 August 2024).
44. EU Reg. 679/2016, GDPR. Available online: https://eur-lex.europa.eu/eli/reg/2016/679/oj (accessed on 11 August 2024).
45. Turin Breathe Committee. Available online: https://www.torinorespira.it/ (accessed on 11 August 2024).
46. City of Turin, Informa Ambiente, Report 26 March 2020. Available online: http://www.comune.torino.it/ambiente/aria/aria_torino/quali-sono-le-cause-dellinquinamento-dellaria.shtml (accessed on 11 August 2024).
47. Air quality in Piedmont. Available online: https://servizi.regione.piemonte.it/catalogo/qualita-dellaria-piemonte (accessed on 5 February 2023).
48. Gallagher, J. A modeling exercise to examine variations of NOx concentrations on adjacent footpaths in a street canyon: The importance of accounting for wind conditions and fleet composition. *Sci. Total Environ.* **2016**, *550*, 1065–1074. [CrossRef]

49. Lin, C.; Ceburnis, D.; Xu, W.; Heffernan, E.; Hellebust, S.; Gallagher, J.; Huang, R.-J.; O'Dowd, C.; Ovadnevaite, J. The impact of traffic on air quality in Ireland: Insights from the simultaneous kerbside and suburban monitoring of submicron aerosols. *Atmospheric Meas. Technol.* **2020**, *20*, 10513–10529. [CrossRef]
50. Gallagher, J.; Gill, L.W.; McNabola, A. The passive control of air pollution exposure in Dublin, Ireland: A combined measurement and modelling case study. *Sci. Total Environ.* **2013**, *458–460*, 331–343. [CrossRef] [PubMed]
51. Ireland's environment: Air. Available online: https://www.epa.ie/our-services/monitoring--assessment/assessment/irelands-environment/air/ (accessed on 1 March 2023).
52. Riondato, E.; Pilla, F.; Basu, A.S.; Basu, B. Investigating the effect of trees on urban quality in Dublin by combining air monitoring with i-Tree Eco model. *Sustain. Cities Soc.* **2020**, *61*, 102356. [CrossRef]
53. Gray, S.C.; Edwards, S.E.; Schultz, B.D.; Miranda, M.L. Assessing the impact of race, social factors and air pollution on birth outcomes: A population-based study. *Environ. Health* **2014**, *13*, 4. [CrossRef]
54. Ferreira, S.; Gallagher, L. Protest responses and community attitudes toward accepting compensation to host waste disposal infrastructure. *Land Use Policy* **2010**, *27*, 638–652. [CrossRef]
55. De Titto, E.; Savino, A. Environmental and health risks related to waste incineration. *Waste Manag. Res. J. A Sustain. Circ. Econ.* **2019**, *37*, 976–986. [CrossRef] [PubMed]
56. CIVITAS Handshake. Available online: https://handshakecycling.eu/turin (accessed on 11 August 2024).
57. Bike Sharing in Dublin. Available online: https://www.dublincity.ie/residential/transportation/active-travel/initiatives/bike-sharing (accessed on 1 March 2023).
58. Sustainability Mobility Plan Introduction. Available online: http://geoportale.comune.torino.it/web/sezioni-tematiche/piano-urbano-della-mobilita-sostenibile-introduzione (accessed on 1 February 2023).
59. Forni, E.; Negro, E.; Carlucci, C.; Nasso, A.; Struppek, M. Actions against air pollution in Turin for a healthy and playable city. *Cities Health* **2019**, *3*, 53–58. [CrossRef]
60. Bena, A.; Gandini, M.; Cadum, E.; Procopio, E.; Salamina, G.; Orengia, M.; Farina, E. Risk perception in the population living near the Turin municipal solid waste incineration plant: Survey results before start-up and communication strategies. *BMC Public Health* **2019**, *19*, 483. [CrossRef] [PubMed]
61. Coi, A.; Minichilli, F.; Bustaffa, E.; Carone, S.; Santoro, M.; Bianchi, F.; Cori, L. Risk perception and access to environmental information in four areas in Italy affected by natural or anthropogenic pollution. *Environ. Int.* **2016**, *95*, 8–15. [CrossRef]
62. Bono, R.; Romanazzi, V.; Bellisario, V.; Tassinari, R.; Trucco, G.; Urbino, A.; Cassardo, C.; Siniscalco, C.; Marchetti, P.; Marcon, A. Air pollution, aeroallergens and admissions to pediatric emergency room for respiratory reasons in Turin, northwestern Italy. *BMC Public Health* **2016**, *16*, 722. [CrossRef] [PubMed] [PubMed Central]
63. Marchetti, P.; Miotti, J.; Locatelli, F.; Antonicelli, L.; Baldacci, S.; Battaglia, S.; Bono, R.; Corsico, A.; Gariazzo, C.; Maio, S.; et al. Long-term residential exposure to air pollution and risk of chronic respiratory diseases in Italy: The BIGEPI study. *Sci. Total Environ.* **2023**, *884*, 163802. [CrossRef] [PubMed]
64. Faustini, A.; Stafoggia, M.; Colais, P.; Berti, G.; Bisanti, L.; Cadum, E.; Cernigliaro, A.; Mallone, S.; Scarnato, C.; Forastiere, F. Air pollution and multiple acute respiratory outcomes. *Eur. Respir. J.* **2013**, *42*, 304–313. [CrossRef] [PubMed]
65. De Donno, A.; De Giorgi, M.; Bagordo, F.; Grassi, T.; Idolo, A.; Serio, F.; Ceretti, E.; Feretti, D.; Villarini, M.; Moretti, M.; et al. Health Risk Associated with Exposure to PM_{10} and Benzene in Three Italian Towns. *Int. J. Environ. Res. Public Health* **2018**, *15*, 1672. [CrossRef] [PubMed] [PubMed Central]
66. Natale, P.; Grosa, M.M. Air pollution trends in the town of Turin in the years 1973–1986. *Sci. Total Environ.* **1992**, *120*, 271–280. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

sustainability

MDPI

Article

Autonomous Generation of a Public Transportation Network by an Agent-Based Model: Mutual Enrichment with Knowledge Graphs for Sustainable Urban Mobility

Flann Chambers [1,*], Giovanna Di Marzo Serugendo [1] and Christophe Cruz [2]

[1] Centre Universitaire d'Informatique, University of Geneva, 1227 Carouge, Switzerland; giovanna.dimarzo@unige.ch
[2] Laboratoire Informatique Carnot de Bourgogne, University of Burgundy, 21078 Dijon, France; christophe.cruz@u-bourgogne.fr
* Correspondence: flann.chambers@unige.ch or flann.chambers@gmail.com

Abstract: Sound planning for urban mobility is a key facet of securing a sustainable future for our urban systems, and requires the careful and comprehensive assessment of its components, such as the status of the cities' public transportation network, and how urban planners should invest in developing it. We use agent-based modelling, a tried and true method for such endeavours, for studying the history, planned future works and possible evolution of the tram line network in the Greater Geneva region. We couple these models with knowledge graphs, in a way that both are able to mutually enrich each other. Results show that the information organisation powers of knowledge graphs are highly relevant for effortlessly recounting past events and designing scenarios to be directly incorporated inside the agent-based model. The model features all 5 tram lines from the current real-world network, servicing a total of 15 communes. In turn, the model is capable of replaying past events, predicting future developments and exploring user-defined scenarios. It also harnesses its self-organisation properties to autonomously reconstruct an artificial public transportation network for the region based on two different initial networks, servicing up to 29 communes depending on the scenario. The data gathered from the simulation is effortlessly imported back into the initial knowledge graphs. The artificial networks closely resemble their real-world counterparts and demonstrate the predictive and prescriptive powers of our agent-based model. They constitute valuable assets towards a comprehensive assessment of urban mobility systems, compelling progress for the agent-based modelling field, and a convincing demonstration of its technical capabilities.

Keywords: agent-based modelling; knowledge graph; sustainable urban mobility; public transportation network; self-organisation; complex systems

1. Introduction

The recent evolution of most urban systems can be summarised by the following observation: the amount of activities and residing population is sharply increasing. As their spatial extent keeps on expanding, and the density of activities and built-up areas keeps on increasing inside them, urban systems are facing significant sustainability challenges [1,2]. In particular, urban mobility is a crucial part of an urban system's integrity [2–4]. As the demand increases for mobility inside, from and towards the city's main activity centers, efficient transportation solutions that also align with sustainability goals to reduce carbon emissions are to be developed.

Planning urban mobility in this scope is a compelling and crucial task for ensuring cities become and stay sustainable. This objective requires an astute and exhaustive assessment of past, present and future trajectories and phenomena related to urban mobility for the given case study [5,6]. Thanks to their descriptive, predictive and prescriptive

powers, agent-based models (ABM) are well-suited for providing such guidance [6]. They are able to help policy-makers understand the inner workings of a given complex system, monitor its behaviour, predict upcoming events, and elaborate what-if scenarios for possible future evolution paths of the system [4–12].

Agent-based models require large amounts of data to function adequately [4,13]. As a result, datasets used in input of such models are often issued from many different sources, and may be formatted in various ways. From there, challenges related to the efficient management of these massive quantities of data, and inter-compatibility issues, may arise [2]. Ontologies and knowledge graphs present themselves as eminent solutions for organising and classifying the information contained in this data [14].

Given these contextual elements, we raise the following research questions : first, what are solutions for designing flexible and versatile agent-based models about urban mobility? To this end, we investigate the adequacy of knowledge graphs as one of such solutions. We highlight how well suited they are as a streamlined, easy to use and powerful descriptive, predictive and prescriptive analysis tool when coupled with agent-based models. Then, how are agent-based models able to enrich these initial knowledge graphs, and how do their self-organisation properties manifest in the scope of an urban mobility study?

We structure this paper in the following manner. Section 2 briefly touches upon the key components of the concepts and methods harnessed in our study. Section 3 delves into the state of the art related to those concepts and methods, and their use in conjunction with each other. Section 4 fully details our methods for conducting our study: creating the knowledge graphs, developing the agent-based model and building the pipeline enabling a tight coupling and mutual communication and enrichment of both entities. Section 5 presents our results, and the most important contribution of this paper: the autonomous generation of our studied system's public transportation network by the agent-based model. Section 6 discusses these results and provides closing thoughts.

2. Background Information

In this section, we briefly introduce concepts that are used in the scope of this study: ontologies, knowledge graphs and agent-based models. We detail some of their key characteristics and the reason we employ them in our works, then we describe our case study and the main attributes we wish to represent.

2.1. Ontologies and Knowledge Graphs

Ontologies are a conceptual framework for arranging and classifying knowledge. They represent a powerful assistant for synthetic complex systems to capture the inner workings of their real-world counterpart [14]. The information is organised into abstract classes and its instances, properties, and restrictions and rules are applied to these categories, in order to question the data and reason about this information.

Knowledge graphs (KG) materialise this abstract organisation of knowledge by representing concepts, instances or data in the form of nodes; and relationships, properties, in the form of edges [14,15]. Subject-predicate-object triplets are born from this paradigm, by directly linking two objects in a graph with an expression (usually verbal) which describes the relationship between these objects (such as "is created by", "approves", etc.).

2.2. Agent-Based Modeling and Its Associated Challenges

Agent-based modelling (ABM) consists in reproducing the behaviour of a collection of individual entities called agents, in order to produce phenomena at various, larger scales. These emerging phenomena are the result of the interactions these agents have between themselves and their environment. These self-organising properties and the high level of detail are key attributes of such models, and the simulations they generate [4,16,17].

This modeling technique comes with many challenges [4,13,18], one of them being the need for high amounts of data, for which we already provided a satisfactory solution in previous research [17]. We utilise the same solution for this study, which is detailed in Section 4. Another conundrum resides in finding an efficient way to incorporate external, expert knowledge in a simple manner, in a way which does not require tinkering with the model's source code, with the goal of making the experimentation process easily accessible to policy-makers and other external parties [4,16]. In this paper, we propose an agent-based model which directly incorporates data from a modular knowledge graph, directly solving the aforementioned problem thanks to the properties of knowledge graphs discussed in Section 2.1.

2.3. Our Use Case: The Greater Geneva and Its Public Transportation Network

The Greater Geneva region is made up by the canton of Geneva and the canton of Vaud in Switzerland, and the departments of Ain and Haute-Savoie in France. The region is composed of many cities, also called *communes*, and is densely populated. Its high attractiveness causes a steady demographic growth [17]. As the region's hypercenter and main economic hub, the city of Geneva, and by extension, the canton of Geneva, constitute the point of convergence of the vast majority of daily commutes, which raises a challenge of great amplitude on mobility inside the region.

In light of these challenges, in the past two decades, Geneva's public transportation system has mainly been developed around its tram line network, spanning multiple major axes linking peripheral areas in the Greater Geneva region to the city's hypercenter. Trams are urban trains that allow for high passenger throughput. They are able to accommodate a higher volume of passengers over longer distances than buses, while servicing many more stops along the way than regular trains can. Our agent-based model will focus on this transportation method, while also considering other existing public transportation solutions offered to the residents of the Greater Geneva region.

3. State of the Art

3.1. Agent-Based Modeling of Urban Systems

Agent-based modeling is an advanced modeling technique for depicting complex systems, such as urban territories, and the interactions they encompass. Agent-based models are particularly well adapted to address topics in urban development [1,6,9,10], such as land cover change, mobility and sustainable development. More specifically, Gaube et al., 2013 [1], utilise agent-based modeling to couple commuting patterns with land use, land cover change and residential dynamics. Motieyan et Mesgari, 2018 [12], develop an agent-based model that investigates how policy-makers can orient their urban mobility planning processes towards encouraging commuters to combine soft transportation modes such as walking and cycling, with public transportation solutions for covering larger distances, instead of resorting to private motorised transportation. The authors show that such planning fosters the sustainable management of urban land use and land cover changes. Other examples in the associated literature, among many others, are the use of ABM for analysing mobility and commuting patterns [6], linking urban planning with the degree of accessibility of basic services to the population inside an urban system [9], or managing the balance between the need for proximity of a bustling, central urban complex and leisurely green open areas [10]. In previous work, we also devised an agent-based model to represent the relocation dynamics in the canton of Geneva, based on the DPSIR framework and actual travel time by public transportation data [17]. Furthermore, agent-based modeling has been shown to be capable of putting together knowledge from stakeholders into one model which is capable of providing guidance for policy-making [2].

3.2. Coupling Agent-Based Modeling and Knowledge Graphs

The idea of coupling agent-based modeling and knowledge graphs stems from the necessity of arranging knowledge into logically organised representations for this specific

modeling technique [16]. The agent-based models' inner structure is compatible with such organisation (classes, instances, properties and relationships [14]), and are thus capable of transcending the knowledge gained from research paper collection.

Agent-based models are capable of putting together knowledge from stakeholders into one model which is capable of providing guidance for policy-making [2]. Gary An [16] recognizes the importance of "knowledge representation" and organisation techniques to advance agent-based modelling research.

While Yao et al. [19] discuss the implementation of a KG-driven ABM, Zafeiropoulos et al. [20] declare ABM as a "valuable tool" for research on socio-ecological systems in general. They develop a KG representing Sustainable Development Goals and its related concepts, and propose a proof of concept implementation of a coupled KG- ABM study.

From our survey of the related literature, we conclude that coupling agent-based modeling and knowledge graphs is an innovative and eminent axis of research, and will dedicate the rest of this paper to detailing a methodology for advancing this field. Our results constitute original research and groundbreaking progress in this direction.

4. Materials and Methods

4.1. Building Descriptive, Predictive and Prescriptive Knowledge Graphs

We mobilise the information organisation powers of knowledge graphs to methodically recount tram line re-structuring events (such as the creation and extension of lines) that occurred throughout the 2003–2024 time period. Each event is decomposed in three major entities: a central **law** which prompted the creation of an investment **credit**, which then planned and enabled the creation or the extension of its associated **tram line**. Each sub-event is dated and organised in a table of events, which is then harnessed to create knowledge graphs of such events. It is of course possible to include events which are planned, but not yet completed (such as the planned extension of a tram line to a neighboring commune), or even hypothetical events at the user's own volition.

In our design process, the produced knowledge graph is in line with the DPSIR framework, a framework which organises key processes of a complex system into five different entities: **D**riving forces, **P**ressures, **S**tate, **I**mpacts and **R**esponses [21]. In previous research [17], we already illustrated how urban processes such as commutes by public transportation and residential choices can be inscribed in a DPSIR analysis graph for building an agent-based model.

Three types of analysis are achieved thanks to this methodology: **descriptive**, by organising and detailing the real history of the tram network; **predictive**, by including planned extensions by policy-makers into the event table, and **prescriptive**, by enabling the user to formulate their own scenarios in an efficient and streamlined manner.

4.2. Coupling Knowledge Graphs and the Agent-Based Model: A Mutual Enrichment Process

The knowledge graphs represent an organised information base that is an invaluable aid in building a flexible and versatile agent-based model. In our study, each knowledge graph represents a particular set of main structuring events which will shape the simulations derived from the associated model. As such, our implementation of these knowledge graphs is able to provide this information directly to the agent-based model in a streamlined manner. Thus, the model is enriched by the knowledge graphs.

Furthermore, this same implementation is able to incorporate outputs generated by the model agents' self-organisation properties back into the original knowledge graph, ensuring a two-way and mutual enrichment of knowledge graphs and the agent-based model. In practice, tram lines and credits generated during the simulation—following the evolution rules described at Section 4.3—are integrated into the graph, sharing a common origin node: the centralized government agent, which approves the creation of an investment credit for all candidate communes.

4.3. Agent-Based Model Evolution Rules

The model follows the evolution rules described in Algorithms 1–6. In all of the algorithms, the variables x_i denote user-defined parameters. The default value is written down in parentheses next to the parameter when possible, or explicitly mentioned after the corresponding formula. Actions are collections of commands that execute when called by a specific agent. Reflexes are actions that are executed at each cycle of the simulations. A cycle corresponds to the execution of all reflexes in the model. The simulation proceeds cycle by cycle until a stop condition is reached. In our model, a cycle corresponds to one day. A concise summary of all evolution rules follows.

At initialisation, all agent types are created: population cells which pave the virtual world, communes, the initial tram line network, roads, and the government agent. The information contained in the knowledge graph is then imported: the policy observed by the communes, the laws, credits and new tram lines. Each of these entities has one or more dates during which they are voted, enacted, created, etc. Their status changes whenever the simulation current date reaches these important dates. Roads are given a weight that depends on their length and the number of people they service within a given radius.

Communes are one of the most important types of agents. Every cycle, if they meet the required conditions, they are able to send an investment credit request to the government, using a clustering and Dijkstra path-finding algorithm which determines the highest-scoring suitable path for a new tram line. The path score p_{score} depends on the length of the path, and the amount of population reached by the path within a given radius. If their credit request is approved by the government, they validate the credit and plan the creation of a new tram line along the path they have found. If the commune is already serviced by a tram line, they create new homes in the cells that are within a given radius of the existing investment credits' spatial extent or tram lines. This action changes (usually decreases) the expansion score of the commune, which is entirely different from the score of a path p_{score}. The expansion score of a commune measures its propensity to grow its population and construct new buildings, and is influenced by the establishment of new homes and the expansion of its neighboring communes.

The government agent is the central power which approves or denies investment credit requests for the communes. It symbolises all political and policy-making powers at play in the real-world system. At regular intervals, the government holds a meeting and decides on the approval of the communes' requested investment credits. If all such credits' score is below a dynamic threshold, no permission is granted and the score threshold and decision interval change (usually decrease). The communes will then have to submit new proposals. Otherwise, the credit is granted to the highest scoring project among all communes, the score threshold is set to the highest value ever seen by the government in this simulation, and the decision interval is reset to its original value.

Finally, the world agent also has its own reflexes. The world turns on prediction mode at a given date, which corresponds to when communes are allowed to start proposing investment credit projects to the government. The world also welcomes a given number of new people each day, which corresponds to a regular demographic growth in the Greater Geneva region, and houses them whenever possible (i.e., when there are empty homes that have been built by the communes).

Algorithm 1 Initialising the agent-based simulation.

Create cells, import amount of people in the cell from shapefile.
Create communes, import name and country from shapefile.
Create tram lines to match the initial network, import line number from shapefile.
Create Roads.
Create Government.

Algorithm 1 *Cont.*

Import Knowledge Graph (KG):

- For communes, set policy to match policy contained in KG.
- Create laws contained in KG, set voting and enforcement date accordingly. Laws will be voted and enforced when the current date of the simulation current_date reaches the voting and enforcement dates. Expansion becomes legalised when one law has been enforced.
- Create credits contained in KG. Set their name, parent law and creation date accordingly. Credits will be created when current_date reaches the creation date.
- Create tram lines contained in KG. Set their name, line number, establishment date and parent credit accordingly. They start white, and will be created when current_date reaches the establishment date, which will then turn them into the color corresponding to their line number.

for each commune, find its immediate neighbors.
Create a weighted graphs of roads for the path-finding algorithm:

- Assign a weight to each road. Road length is divided by $\left(1 + x_1 \frac{n_{people}}{n_{people,max}}\right)$, where x_1 defaults at 1, n_{people} is the amount of people within a x_2 (400) meter radius around the road, and $n_{people,max}$ is the maximum value of n_{people} across all roads. Therefore, the more people a road reaches, the shorter it is made virtually for the Dijkstra path-finding algorithm.

Communes which contain tram lines or credits in their area **do** "declare target cells for construction". ▷ See Algorithm 3

Algorithm 2 Cells agent actions and reflexes.

(0) Initialisation

- Amount of people $nPeopleAround \leftarrow$ total amount of people in a x_{s1} (500) meter radius around the cell, including people in the cell itself.

(1) Action "update status"

- Status becomes "near tram line", "near credit" or "isolated", depending on the presence of tram lines and credits at distance x_{s2} (400) meters or less, with priority given to tram lines.

Algorithm 3 Communes agent actions.

(1) Action "Create a tram project"

- Enter project cooldown (will exit cooldown when the government makes a decision).
- Generate paths:
  - Cluster the x_{c1}% (10%) highest-populated cells in the commune, based on their $nPeopleAround$ value.
  - **for** each cluster **do**
    - Create a path from each cell in the cluster to the closest tram line, and compute its path score p_{score}. $p_{score} = x_{c2} \frac{(n_{people})^{x_{c3}}}{(l)^{x_{c4}}}$, where n_{people} is the amount of people within a x_{c5} (500) meter radius from the path, l is the geometric length of the path in meters, x_{c2} defaults at 1'000'000, and x_{c3} and x_{c4} default at 1.
- Among all of the specific commune's generated paths, submit path with the highest p_{score} to the government:
  - Ask government to include commune name to its queue.
  - Delete all non-highest-score paths generated by the commune.
- The commune is put on standby, awaiting a credit approval response by the government.

Algorithm 3 *Cont.*

(2) Action "Generate a new credit"

- Create a tram line.
  - Shape corresponds to the previously submitted path.
  - A credit is first set to be created at current_date + x_{c6} (6) months.
  - A tram line is then set to be created at date current_date + x_{c7} (24) months.
  - Newly created credits and tram lines are exported to be integrated back into the original KG.
- Ask cells to update their status.
- Remove the submitted path from the model.
- Declare target cells for construction: Set target cells to the set of non-full cells at a distance x_{c8} (400) meters or less from all credits and lines inside the commune's area.

(3) Action "Build n new homes on target cells."

- Target cell c **does**:
  - Increase $nNewHomes$ by n.
  - Increase $nEmptyHomes$ by n.
  - **if** $nNewHomes$ exceeds the maximum amount of homes to build for this cell x_{c9} (100) **then**
    - Flag self as full.
    - Remove self from the commune's target cells for construction.
- Increase total amount of empty homes by n.
- Add $n \cdot x_{c10}(-3)$ to expansion score of the target cell c's commune.

Algorithm 4 Communes agent reflexes.

(A) Main reflex

- The following conditions must be fulfilled :
  - Prediction mode is on.
  - Expansion has been legalised by the previous enforcement of a law.
  - The commune's policy allows expansion.
  - The expansion score is above the expansion score threshold x_{c11} (10).
  - The commune is not awaiting a credit approval response from the government.
  - There is at least one neighboring commune which is linked to the tram line network.
- **if** there are no tram lines in the commune's area **then**
  - If not in project cooldown, the city center is determined by selecting the cell in the commune with the highest amount of people within a x_{s1} (500) meter radius, then create a tram project.
- **else if** there are target cells for construction **then**
  - Build 1 home on a random target cell.

Algorithm 5 Government agent actions and reflexes.

(1) Approve credit

- Multiply score threshold Ωp_{score} by $\left(1 - x_{g1}\frac{n_{new}}{n_{tot}}\right)$, where n_{new} is the amount of new people, n_{tot} is the amount of people in the system, excluding new people, and x_{g1} defaults at 2.
- Set target commune for approval to the commune with the submitted path with the highest path score p_{score}.
- **if** the target commune submitted path score p_{score} is higher than the threshold Ωp_{score} **then**
  - Score threshold Ωp_{score} is set to the maximum path score p_{score} ever authorized by the government.
  - Decision interval Δt_d is reset to its original value x_{g2} ($3 \cdot 366$ days).
  - Clear queue.
  - Target commune generates new credit.

Algorithm 5 *Cont.*

- **else**
  - Score threshold Ωp_{score} is multiplied by x_{g3} (0.8).
  - Decision interval Δt_d is multiplied by x_{g4} (0.9), and set to 31 if below 31.
- Communes become no longer in project cooldown and awaiting a credit approval response.

(A) Main reflex

- Happens only when current_date is equal to the next decision date.
- **Do** approve credit.
- Set next decision date to current_date + decision interval Δt_d.

Algorithm 6 World agent actions and reflexes.

(A) Main reflex

- **if** past the start population growth date **then**
  - Increase population (add x_{w1} (18) new people to the amount of *newPeople*: $nNewPeople \leftarrow nNewPeople + x_{w1}$).
- **if** there are new people ($nNewPeople > 0$) **then** house new people, **while** there are remaining new people and empty homes **do**:
  - Housing cell *hc* <- one of cells with empty homes, with priority given to cells near a tram line.
  - Ask housing cell *hc* to reduce its amount of empty homes by 1, and its amount of housed people by 1.
  - Reduce the amount of new people by 1.
  - Add x_{w2} (1) to the housing cell's commune's expansion score.
  - Add $\frac{x_{w3}(1)}{\text{amount of neighbors}}$ to the neighboring communes' expansion score.
  - Decrease total amount of empty homes and total amount of new people by 1.
- **if** date of prediction start has been reached **then**
  - Prediction mode is turned on.
- **if** date of end simulation has been reached **then**
  - Export outputs and pause the simulation

4.4. Data Sources

For this study, data sources are split into two groups: one for elaborating a knowledge graph which adheres to real-life public tram network re-structuring events as they unfolded during the 2003-2024 period in Geneva, and another group for building the agent-based model.

Elaborating the first event table required the meticulous examination of a variety of sources—the lack of academic research putting down to paper such history urges us to consider other sources. The information contained in articles from Wikipedia, the local press, related project information websites and policies enacted in Switzerland, are carefully and critically reviewed and cross-examined by the authors based on their collective knowledge of the history of the area [22–37]. The event table devised based on this information is presented on Table 1, and the corresponding map is presented on Figure 1. The knowledge graphs requires the exact date of each event, with a day, month and year. When the information on the day or the month is missing, the default value is set to the first day of the month, or the first month of the year.

The data used for building the agent-based model originates from the following sources: the SITG (***S**ystème d'**I**nformation du **T**erritoire **G**enevois*, https://ge.ch/sitg/, accessed on 21 August 2024) web platform for open geomatic data on the canton of Geneva and the Greater Geneva region, and the IGN (***I**nstitut **G**éographique **N**ational*, https://geoservices.

ign.fr/route500, accessed on 21 August 2024) web portal for geographic data in France. Table 2 summarizes the statistical and geographical data harnessed from these sources.

Table 1. Summary of all Geneva tram network re-structuring events since 2003, which is then incoroporated into the knowledge graph.

Tram Line Number	Name of the Tram Section	Itinerary Known in (Credit Established)	Line Inauguration
14	TCMC Avanchet	2003	07.12.2007
14	TCMC Meyrin-Graviere	2003	12.12.2009
18	TCMC section CERN	2003	02.05.2011
14	TCOB Bernex P+R	2006	11.12.2011
14	TCOB Berney-Vailly	03.2010	04.07.2021
17	Extension Annemasse	09.2013	14.12.2019
15	Extension Saint-Julien—section ZIPLO	02.12.2016	10.12.2023
15	Extension Saint-Julien—final section	02.12.2016	2027
17	Extension Le Perrier	2020	2026
13	Ferney-Voltaire	10.2020	12.2026

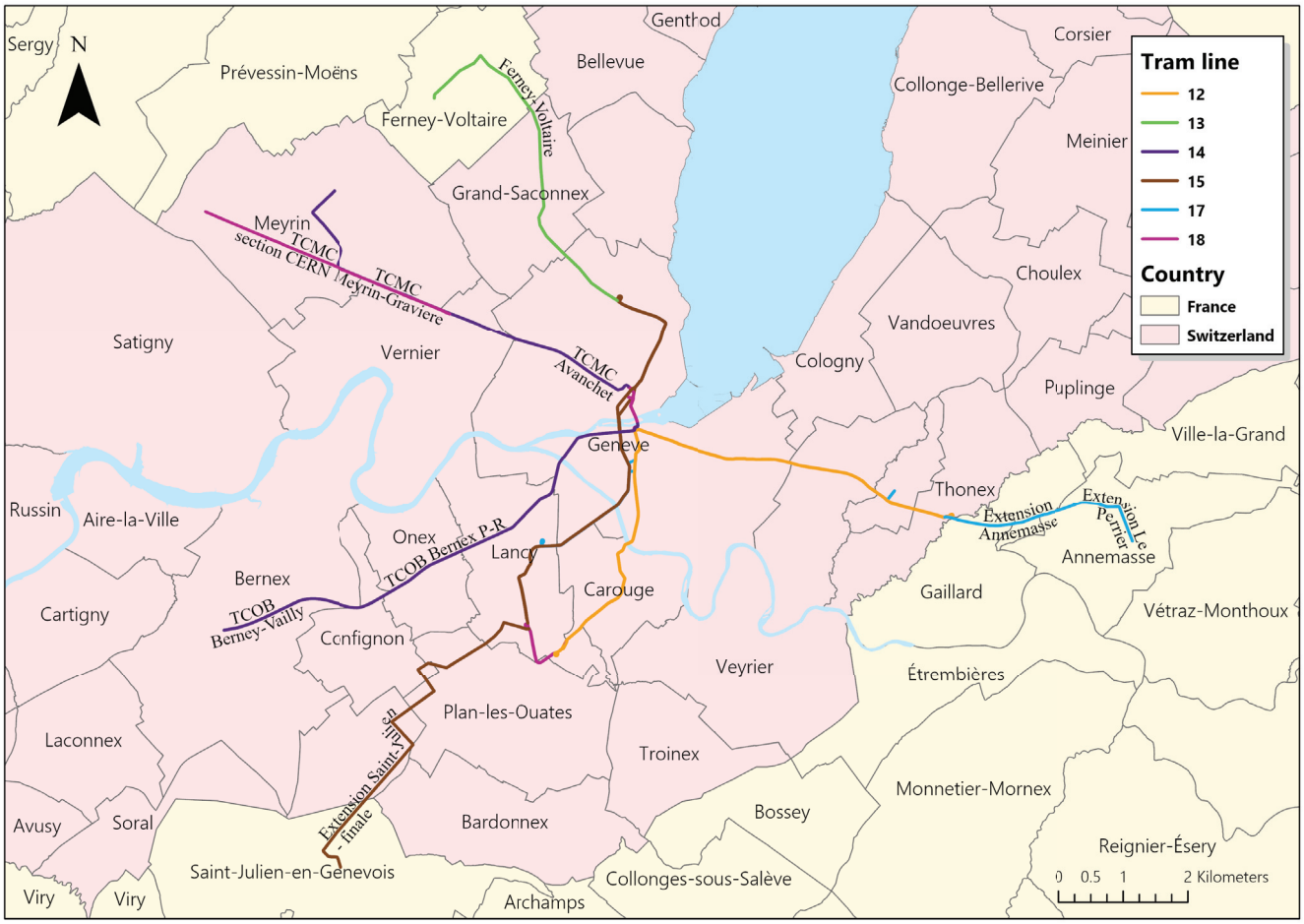


Figure 1. Map of the Geneva public transportation network's tram lines. Present date and future segments are represented according to the event table devised in this study (see Table 1). Tram line 13 (towards Ferney-Voltaire), tram line 15 extension to Saint-Julien from ZIPLO, and tram 17 extension to Le Perrier from Annemasse center, are planned works as of 2024, and are not yet established.

Table 2. List of datasets used for building the agent-based model.

Source	Dataset	Raw Dataset Name	Spatial Features	Attributes
SITG	Population cells	AGGLO_CARREAU_200	✓	Population in 2019
SITG	Communes	AGGLO_COMMUNES	✓	Name, country
SITG	Roads	GMO_GRAPHE_ROUTIER	✓	
IGN	Roads	ROUTE500 (2021-11-03)	✓	
SITG	Tram lines	TPG_LIGNES	✓	Name, line number
SITG	Lake	GEO_LAC	✓	Name

4.5. Workflow and Implementation

The workflow for the implementation of the methods described previously is displayed on Figure 2. The remainder of this section is broken down into exhaustive descriptions of each step.

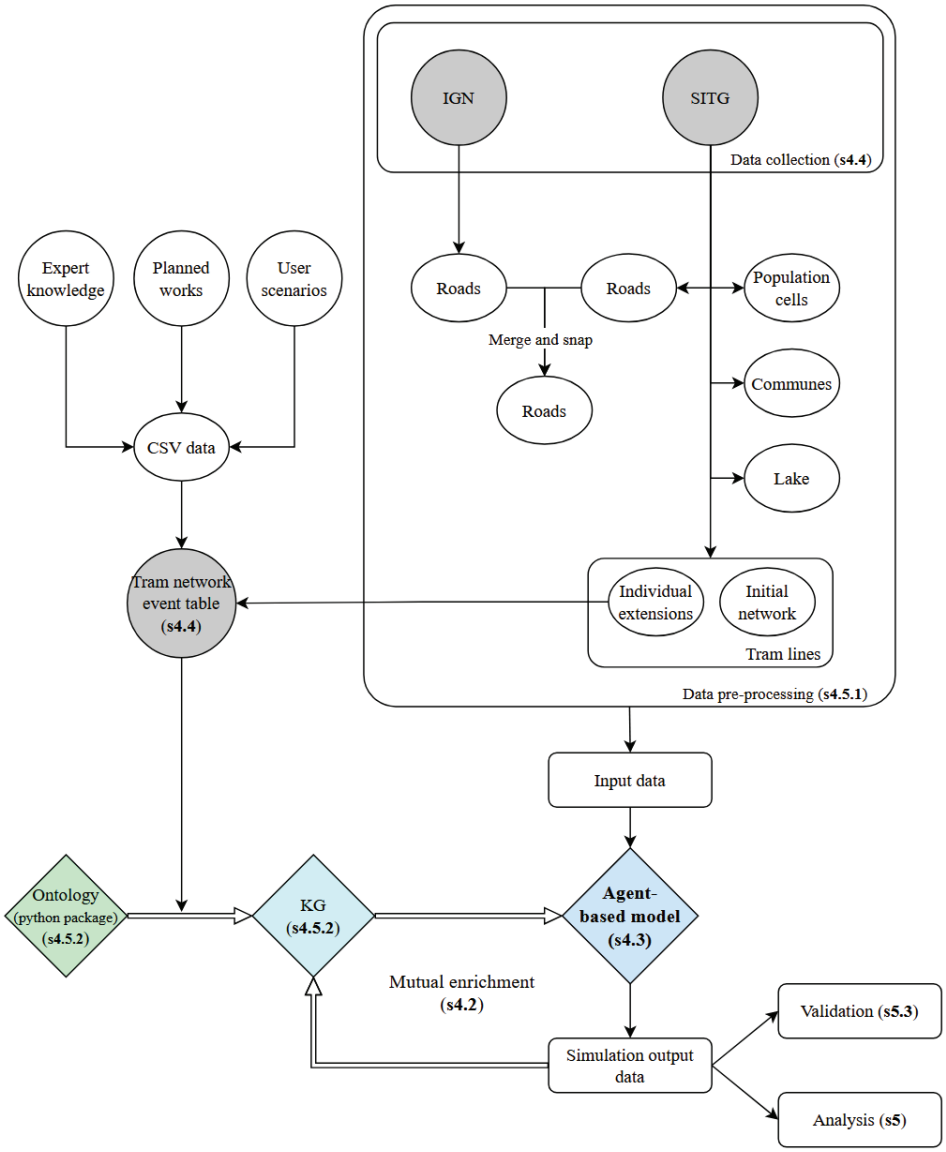


Figure 2. Methodology workflow for gathering input data, building the knowledge graph and agent-based model evolution rules, producing simulation output data, then visualising and analysing it, and finally validating the model. When relevant, elements of the workflow are indexed with their dedicated section in this paper, denoted as (**s4.1**) for Section 4.1, for instance.

4.5.1. Data Pre-Processing

For the road network, two different shapefiles were downloaded, one for each country over which the Greater Geneva region spans (Switzerland and France). The two networks were merged together thanks to the *Merge* and *Snap* tools of the proprietary GIS software ArcGIS Pro 2.4.0, although any alternatives, such as QGIS, may accomplish the same task.

Tram lines were isolated from the full TPG transportation network thanks to the *Select by Attributes* tool. Then, each section described in Table 1 was isolated and saved as an independent shapefile. The remaining network constitutes the initial network loaded by the simulations in the first place. Tram segments that do not yet show in the SITG dataset because they are planned works (such as tram line 13) are drawn by hand based on the information contained in the sources for building the event table, and saved as independent shapefiles. Finally, to complete the tram line pre-processing, another column is added to the event table, which contains the name of the associated shapefile for each tram section. This name will be vital for the knowledge graph to access the correct shapefile when processing the event table into a format that is readable by the agent-based model at simulation initialisation.

4.5.2. Ontology and Knowledge Graphs

We develop a python package, which structures input information according to the ontology and produces the knowledge graphs. It is fully equipped with the necessary tools to gradually build knowledge graphs of public transportation network modification events, import them from a CSV-format table, as well as visualise and export these graphs to the associated agent-based model. The package is also able to import the aforementioned model's outputs into the original knowledge graph, constituting the enrichment of such graphs.

Figure 3 shows the ontology's structure, with the original elements in white and the information subsequently imported from the agent-based simulations in light blue. The initial ontology follows the law $\longrightarrow$ investment credit $\longrightarrow$ tram line logic mentioned in Section 4.1 and is also equipped with the expansion policy of each commune. The information imported from the simulations are organised in the following manner: a centralised **government agent** authorises a given **commune** to create an **investment credit**, which will later establish a new **tram line** or a tram line extension.

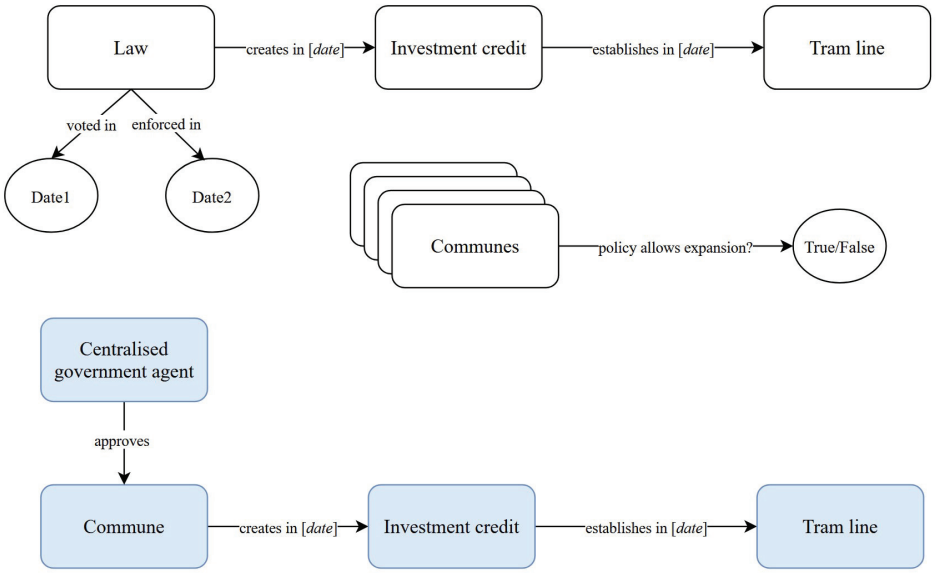


Figure 3. Organisation of elements inside the ontology. White boxes show elements that are contained in the ontology used to build the initial knowledge graphs, which then enrich the agent-based model. Blue boxes show elements that are imported from the agent-based simulations, which then enrich the original knowledge graph.

The graph is built by calling a small amount of simple functions from the python package. We choose the jupyter notebook format to provide a template, which also serves as a tutorial for the user to start building their own knowledge graphs in a few guided steps. The method contained in the template is described in Algorithm 7. The template can be cloned at will, and each copy can be individually filled in for any set of past events, future plans or hypothetical scenarios. Then, making the agent-based model adhere to a particular storyboard is as easy as running the *.ipynb* notebook file of choice prior to running the agent-based simulation.

Algorithm 7 Template for building knowledge graphs of the public transportation network's expansion events.

Import the ontology package:

```
from ontology import *
```

Create a new law, then vote and enforce it at given dates.

```
a = Law('lawname')
a.vote('YYYY-MM-DD')
a.enforce('YYYY-MM-DD')
```

Import a list of credits and tram lines from an event table. The table should also contain, for each event, the name of the associated shapefile which contains the final spatial extent of the tram line.

```
importEventTable('table.csv')
```

Import communes' name and expansion policy (i.e., welcome a new tram line, construct new residential buildings).

```
importCommunes('communes.csv')
```

Change expansion policy for a given commune.

```
getCommune('Geneve').changePolicy()
```

Display the built knowledge graph.

```
showKG()
```

Export the knowledge graph as one of the agent-based model's input.

```
exportKG()
```

Import the agent-based model's outputs back into the knowledge graph.

```
importModelOutputs('folder_name')
```

4.5.3. Agent-Based Modelling in GAMA

We implement the agent-based model using the GAMA platform [38]. This choice was motivated by the platform's ability to produce powerful spatially-explicit simulations in a versatile manner [4]. The model is carefully crafted to directly integrate the information contained inside the knowledge graph built at the previous step. As such, it is mandatory to verify that the correct knowledge graph has been generated, as described in Section 4.5.2, before running any simulations.

The model implements all behavioral rules described at Section 4.3. The model is able to save agent attributes as well as a list of all newly created investment credits and tram lines, either by pressing a button during the simulation or automatically at the defined date of the simulation's end. The list of attributes and information exported by the model is detailed in Table 3.

Table 3. The agent-based model's exported attributes for the tram lines, credits and cells agents.

Tram Lines	Investment Credits	Cells
Name	Name	Amount of people
Line number	Requester commune name	Amount of people around
Date built	Date created	Amount of new homes
Parent credit name	Child line name	Amount of empty homes
Built? (yes/no)	Created? (yes/no)	

4.5.4. Output Data Processing

Apart from being integrated back into the original knowledge graph, the model's outputs, such as the spatial extent of the newly created tram lines, and agent attributes (shown in Table 3), are processed in the proprietary GIS data analysis and visualisation software ArcGIS Pro. We are thus able to present detailed maps recounting the history of the tram line network's evolution in the agent-based model.

4.6. Validation

The validation process of our model will be carried out in two different steps. First of all, the model harnesses data from the real Greater Geneva system it focuses on. Indeed, all shapefiles originate either from the SITG portal for geomatic data about the canton of Geneva, or the IGN portal for geographic information on the French territory. The simulations derived from the model incorporate the spatial positions and various attributes of all the objects it uses. These design choices enhance our model's conformity with the real system's functioning.

Moreover, the model will be tasked to recreate the actual present-date real world public transportation network from a blank state, and will be judged based on the results of this endeavour. The map of the simulated network will be compared with the current TPG network. This experiment will not only serve as validation, but also as a technical experiment that we believe is of great importance and interest for the agent-based modelling field applied to urban mobility.

5. Results and Discussion

The results of our study include three simulation output datasets, for which all of the user-defined parameters are set to their default value indicated in Algorithms 1–6. The first simulation narrates the case study's actual history and displays the planned tram line extensions for the years to come (real-world data and current situation). The next two simulations strive to recreate the Greater Geneva's tram line network from a blank state.

5.1. A Model Capable of Replaying Past Events, Predicting Future Events and Exploring User-Defined Scenarios

The first simulation incorporates the knowledge graph built from the event table pictured on Table 1 (real-world data and current situation). Figure 4 displays the map of the network obtained at the end of the simulation, with each new tram section named and dated. Figure 5 shows the final knowledge graph obtained after enrichment by the simulation outputs.

Thanks to the simulation, we are able to obtain a short film of the evolution of the tram network, that transcends the textual data at its origin. The current real-world network is composed of 5 different tram lines and services 15 different communes. A sixth tram line is created as part as the network's planned future developments, which plan to service 4 additional communes. We also observe extensions that were not planned by the event table, but were rather created by the artificial system and its self-organisation properties, based on the spatial distribution of the population and the road network. The first of such extensions is one to the center of Satigny, then follow extensions towards Saint-Genis-Pouilly, and the neighboring communes Thoiry, Saint-Jean-de-Gonville and Crozet, for a total of 5

additional communes reached by the final network. The first two extensions can easily be explained by the presence of a train line directly linking Satigny to the center of Geneva, and the former plan to extend tram line 18 to Saint-Genis-Pouilly, which was ultimately scrapped. The degree of confidence in the prediction of the tram line 18 extension to Saint-Genis-Pouilly is greatly enhanced by the past elaboration of this plan by local policy-makers. The next extensions rely on optimistic assumptions that lies in the continuity of the works undertaken along the tram line 18 axis, and in order to see these predictions become a reality, local policy-makers must go through the process of investigating the feasibility, viability and desirability of such works.

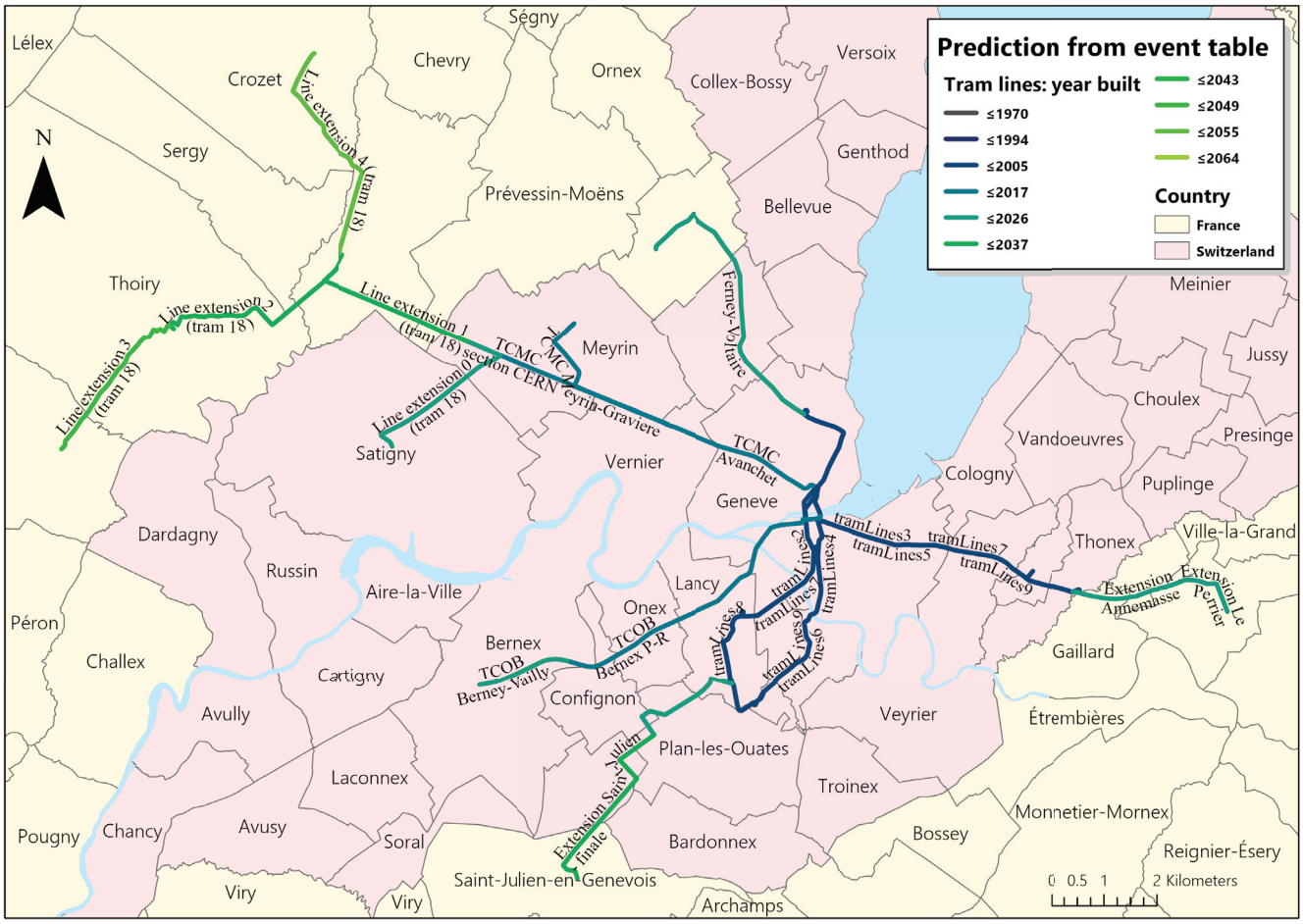


Figure 4. Tram network map for simulation 1, with the initial network obtained from the event table on Table 1 (real-world data and current situation), and prediction mode enabled for the future time-lines.

5.2. A Model Capable of Creating an Artificial Transportation Network from a Blank State

The first simulation showed a preview of the model's predictive and self-organisation properties, now we push these to the extreme by supplying the bare minimum in terms of initial tram network to the model. Only a small loop around the Geneva city center is initially provided to the model, the simulation then builds a whole public transportation network, only on the foundations of the spatial distribution of the population and the existing road network. The initial network is pictured in dark red on Figure 6, which shows the results of the simulation, while Figure 1 reminds us of the actual present-day network.

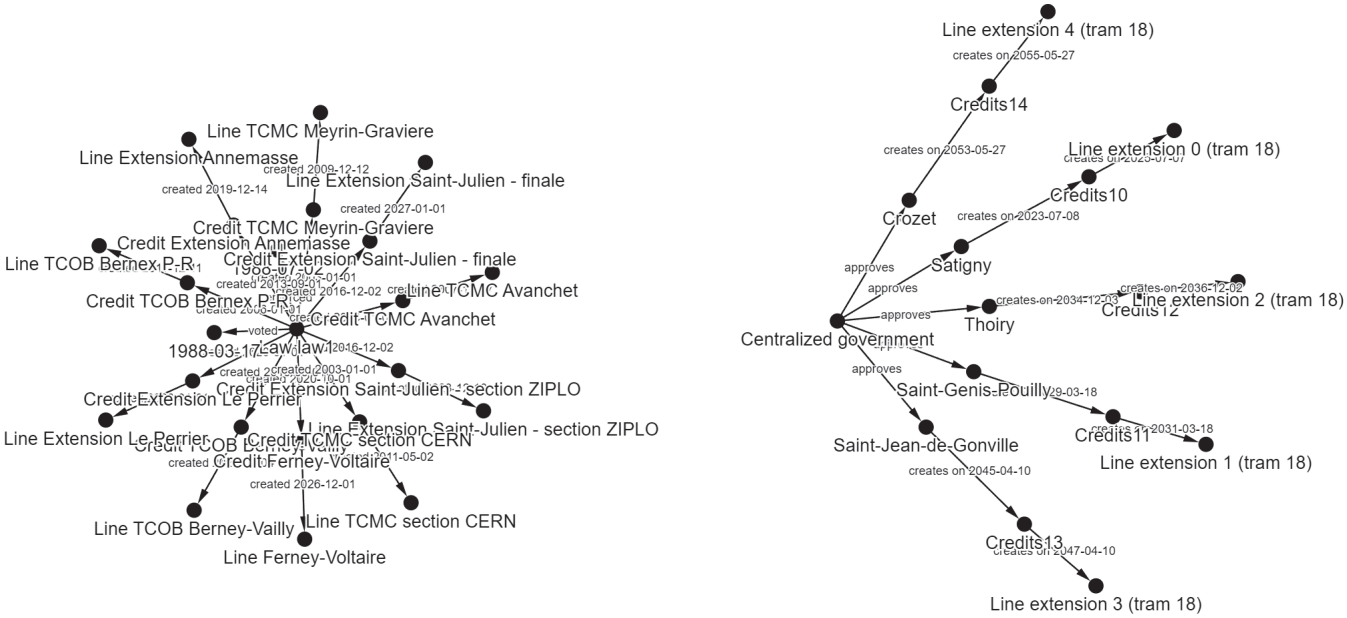


Figure 5. Initial knowledge graph created from the event table on Table 1 (**left**), used as input for the agent-based model, and nodes generated from the simulation outputs (**right**).

The artificial network traverses up to 25 communes, and its overall structure is pleasantly similar to the real-world network. The connection to Meyrin is one of the first axis to be established, and matches most of the itinerary of tram lines 14 and 18 in the real world system. The discrepancies that can be observed near the city center of Geneva can be explained by the fact the real-world tram lines 14 and 18 are designed to call at the Cornavin station, and that the road between Geneva and Vernier is a high-priority axis for public transportation, that is already serviced by up to four bus lines. In the timeline, the loop around Lancy and Carouge closely follows, which corresponds to the loop formed by the actual tram lines 12, 15, 17 and 18. Then, extensions to Bernex (real-world tram line 14), Satigny (train), Saint-Genis-Pouilly and Thoiry are established, which match the results observed in the first simulation. The eastern part of the network, towards the French communes of Gaillard and Annemasse, are developed much later, by implementing small extensions for each commune lying between these communes and the Geneva city center. This is owing to the fact the tram line 12, which services these regions, is the most ancient tram line in Europe. It originates from the second half of the 19th century, which is why the timeline for the establishment of the artificial network in this region is less relevant than the fact the model actually recognises this region as an important part of the real-world network.

We notice that tram extensions to Ferney-Voltaire and Saint-Julien-en-Genevois are missing and never built by the model. This is due to a model quirk, which prevents communes who do not own a neighbor that is connected to the network, from requesting an investment credit to be attached to the network. This is the case for Saint-Julien-en-Genevois, and for Ferney-Voltaire for most of the simulation. In the case of Ferney-Voltaire, the only practical road for extending a tram line is the one following the tunnel under the Geneva international airport, which goes through Grand-Saconnex, a commune that was never actually connected to the artificial network. Hence, the model was not able to predict these particular planned works.

Some artefacts can be observed in the form of small sections that are not realistic for a tram line itinerary. The model is indeed not capable of smoothing out such granularity and asserting a long straight-line axis for its artificial tram lines, which is the usual real-life methodology for planning such works. However, the model is capable of declaring which communes are susceptible to request a connection to the tram network, in which order, and

which ones are the most likely to be approved by a centralized government entity, that oversees the transportation network development for the whole Greater Geneva region.

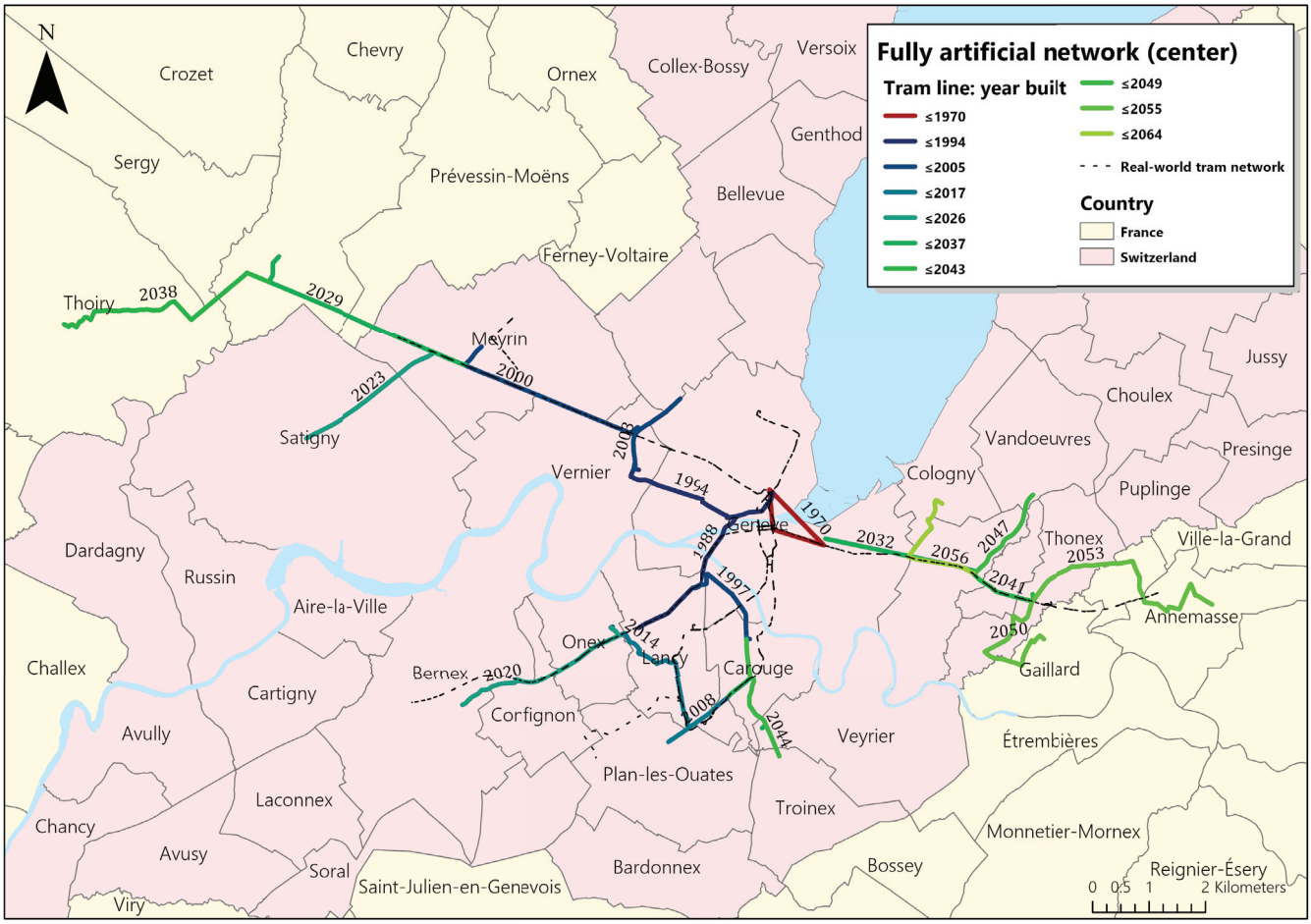


Figure 6. Tram network map for simulation 2: fully artificial network obtained from a small loop in the city center of Geneva (in dark red).

In an effort to better account for the seniority of tram line 12, linking the center of Geneva to the eastern part of the canton, we produced a third simulation, in which we used a small section of the ancient tram line 12 as the base network for the model to build upon. It is pictured in dark red on Figure 7, which shows the results of the last simulation. In this simulation, the transportation network generated by the model reaches up to 29 communes.

The results bear a small set of differences with respect to the previous simulation. The dynamics of the simulation for the establishment of the Lancy-Carouge loop, the extensions towards Bernex, and towards Meyrin, Satigny, Saint-Genis-Pouilly and further, did not change. Part of tram line 12 was recreated early, but the extension towards Gaillard still happened much later than the aforementioned works. Two major changes in the network happen however, one towards Ferney-Voltaire, for which this simulation correctly predicted its connection to the existing tram network, and the other consisting of a major axis along the left bank of the Leman lake towards Collonge-Bellerive and Meinier, for which the equivalent in the real world are bus lines.

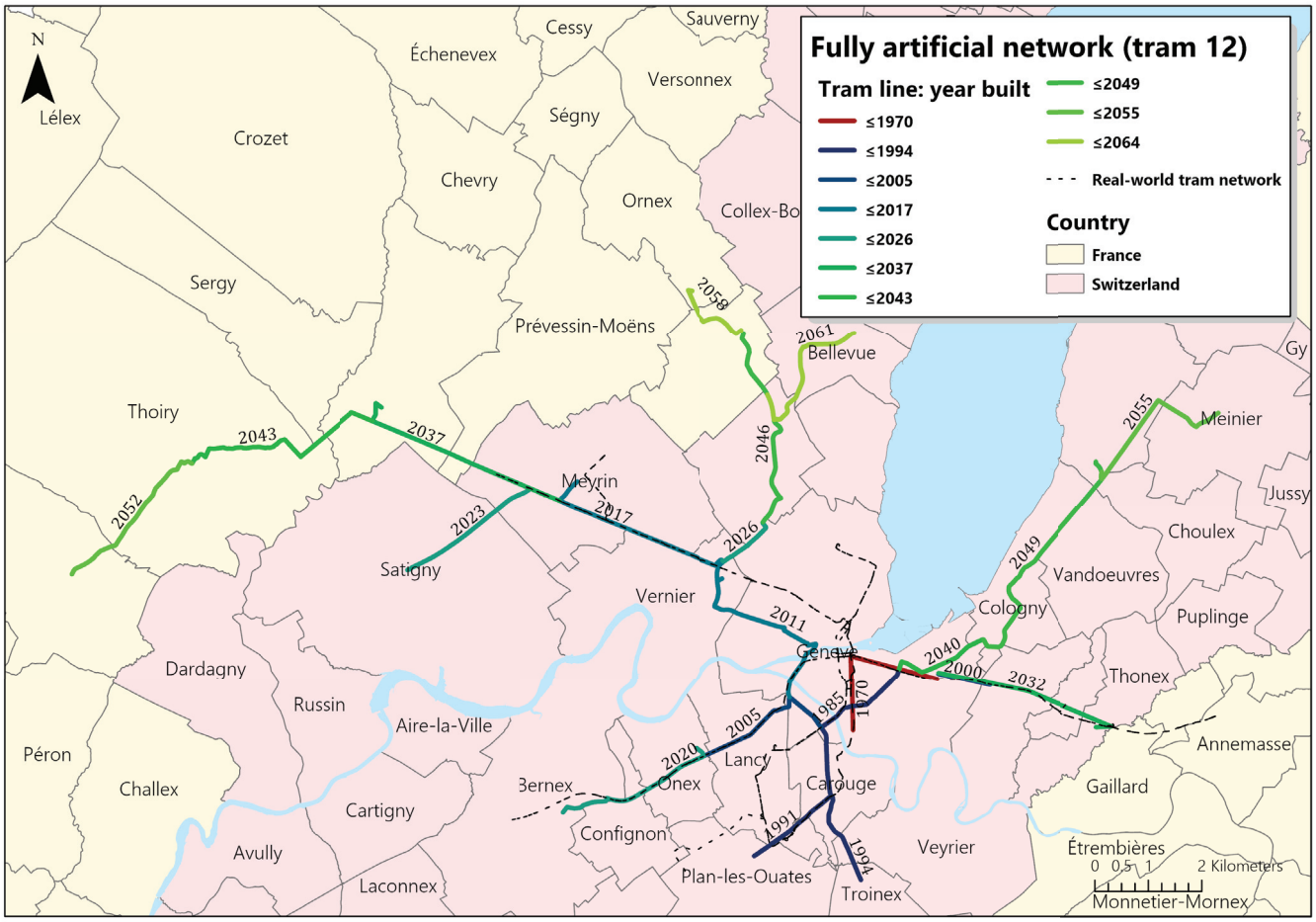


Figure 7. Tram network map for simulation 3: fully artificial network obtained from the initial tram line 12 extent (in dark red).

The only cause of these differences is the alternative geometry of the initial network upon which the model builds the artificial network. The initial value of all the other parameters, and all the evolution rules of the model, did not change. A small change in the network's initial conditions can lead to significant differences in its final configuration, which is an indicator of a complex system. Indeed, in our case, each time a section of the network is built, it takes into consideration the geometry and spatial extent of the current network, and the spatial distribution of the surrounding population. Since the geometry of the network's expansions are highly sensitive to these parameters, a modified initial network may cause the first segment of the generated network to differ as well, propagating the discrepancies to further segments as the network is generated segment-by-segment. This can explain the fact that some communes are connected to the network before others. Nevertheless, and despite the change in dynamics of the simulation caused by differing initial conditions, the main features of the real-world transportation network are re-created by the model, which solidifies its predictive and prescriptive capabilities for our case study.

Going back to the link of our agent-based model with knowledge graphs, both simulations 2 and 3 were tasked to autonomously generate a full transportation network for the Greater Geneva region from a small initial network, and thus did not incorporate information from a knowledge graph. However, they both *produced* a knowledge graph (see Figures 8 and 9) that describes the establishment of the artificially generated tram network, segment by segment, and following the same logic: the central government agent approves the commune's request to create an investment credit, that then is used to create the associated tram line.

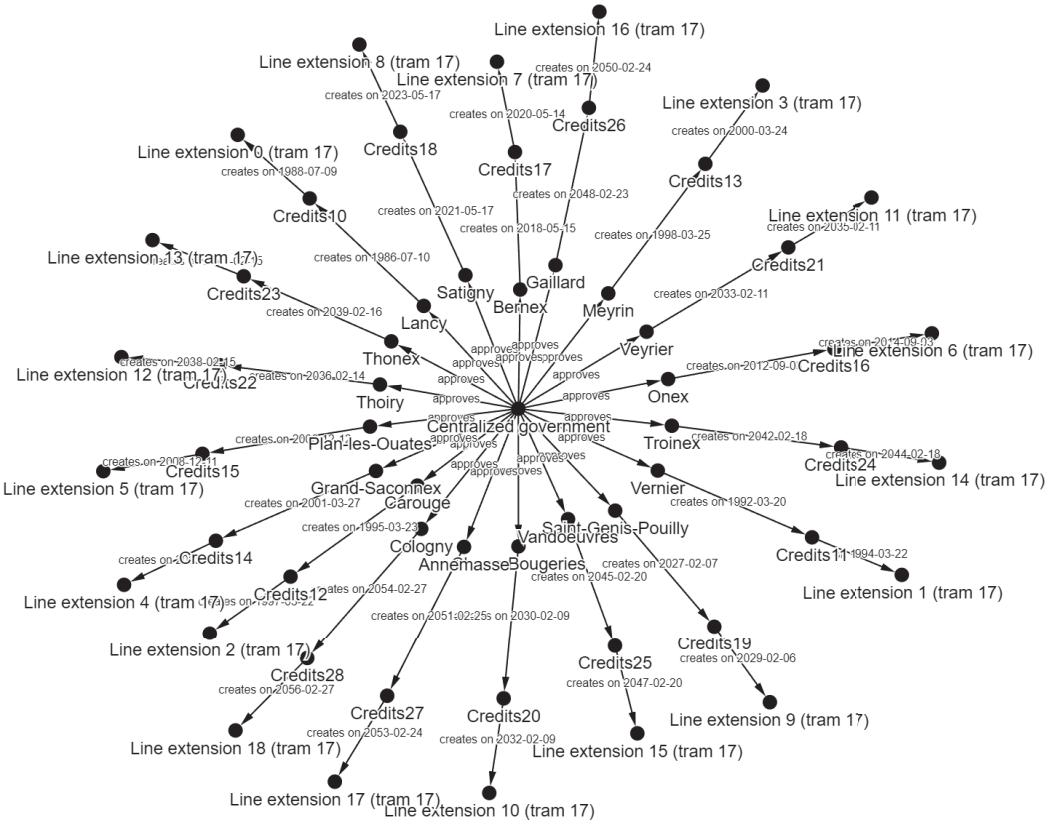


Figure 8. Knowledge graph created from the model outputs during simulation 2: transportation network artificially generated by the model based on a small loop near the city center.

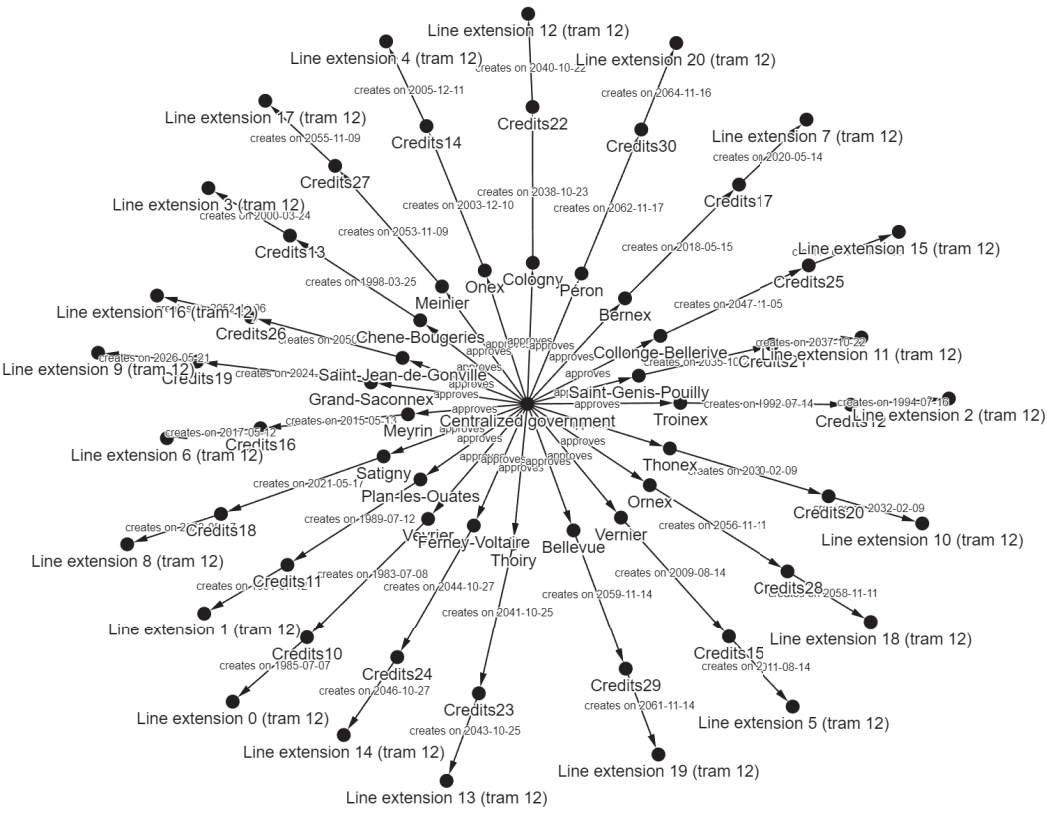


Figure 9. Knowledge graph created from the model outputs during simulation 3: transportation network artificially generated by the model based on the initial tram line 12 extent.

5.3. Validation

As our validation process, our model was tasked to recreate a whole artificial public transportation network for the Greater Geneva region. Such a task was achieved by the model by using only two parameters: the spatial distribution of the population and the existing road network. The results obtained when comparing the artificial networks with the real-world one proved satisfactory. The few differences observed are easy to link to real-world events or quirks in the model's evolution rules. This analysis, coupled with the fact the model's components are based upon real world data, such as geomatic data from the SITG and IGN portals, constitute our model's validation process.

5.4. Shortcomings and Future Prospects

Given the nature of the tram transportation mode, usually requiring long straight axes to be fully efficient, topological constraints need be considered to better the model. One of the shortcomings of the current implementation of our model resides in its inability to capture this facet of tram network development. Additionally, the algorithm for the initial implementation was in fact rather complex to write in GAMA, despite taking into account only two parameters for the generation of a public transportation network. This shortcoming must be solved before more parameters can be integrated inside the model.

From there, to further enhance the credibility of our model, we identify the consideration of more parameters for creating the artificial network as an eminent axis of improvement. These future works are aligned with those from similar studies, which are often centred around incorporating more agent types, improving existing behaviour rules and adding more behaviour diversity among agents and in reaction to various environmental events or other agent groups' behaviour [12,39–47]. To achieve such a goal, the implementation of the various algorithms invoked in this study must be more refined, in order to eliminate related quirks and support the enrichment of the model with more features. In short, optimising model code, smoothing out details and adjusting more parameter values, with the objective of creating simulations that are more faithful to the terrain reality, would improve the validation process and predictive powers of our model.

Finally, transferring the model to other use cases would be a compelling axis of improvement, which is made possible by the high modularity of agent-based models [17]. Use case transferrability is a concept invoked by related studies in their closing discussion [39,41,43] as a valuable strength of the agent-based modelling methodology. It owes to the high parametrisation and versatility offered by this method. Such an endeavour is achieved by replacing existing datasets by those of another area, while keeping the model's initial rules of evolution, or tweaking them to account for the new use case's peculiarities. For instance, to transfer the study to the region around the city of Lausanne (Vaud, Switzerland), the geometry of the virtual world and datasets such as the spatial distribution of the population and public transportation and road networks would be those of the canton of Vaud (instead of Geneva), while the agents would still exhibit the same behaviours as the model for the Greater Geneva region (such as communes generating investment credits and sending tram line establishment projects to be considered by the governing parties). Extending this methodology to other use cases would not only be interesting for the new systems studied, but also for the model itself: if a sound analysis can be provided by the model for additional case studies as well, then its validity is further proven.

6. Summary and Conclusions

The methods and results we discuss in the scope of this study constitute valuable assets for the field of agent-based modelling and knowledge graphs applied to urban mobility. We developed a python package harnessing a knowledge graph for preparing and organising data from a variety of sources to be incorporated directly into our agent-based model. The pipeline ensures tight coupling of the agent-based model and the knowledge graph. It provides a satisfactory solution to one of the agent-based modelling field's sharpest challenges: efficiently managing the wealth of data provided to the model in input. Users

from a wide variety of domains are able to incorporate their very own tram line history version, formulate their own predictions, and test various what-if scenarios, in an easy and streamlined manner, thanks to the clear instructions provided by the jupyter notebook template—no advanced programming skills are needed. This enables a versatile and flexible customisation of the simulations produced by the agent-based model, and unlocks the possibility of automating the production of scenarios and subsequently exploring them with the model. In return, the knowledge graphs are enriched with the information generated by the model in output thanks to its self-organisation properties. The model is capable of predicting future tram line extensions based on the existing network, and those predictions are easily imported back into the original knowledge graph.

Results show that the model is capable of replaying the history of the tram line network in the Greater Geneva region, and predicting future tram line extensions, whether they are registered in the knowledge graph built by the user or generated by the agents' collective behaviour. We tasked the model to recreate the real-world tram network from scratch, starting from the city center of Geneva, and obtained satisfactory results. The model builds tram lines only based on the existing road network and the spatial distribution of the population, aiming to maximise the amount of people reached by the tram lines. The resulting network closely resembles its real-world counterpart, and most of the communes serviced by the actual network are accurately connected in the artificial one. This further strengthens the adequacy of our model in representing complex urban mobility related processes, and demonstrates that the spatial distribution of the population, and working to maximise the amount of people reached by a transportation network, are strong indicators of the efficiency of such networks. Our agent-based model's capabilities in building a sound artificial transportation network are promising results for agent-based modelling studies for sustainable urban mobility. They open up possibilities for new research focusing on other case studies, that examine the artificial networks the model would establish for these case studies and compare them to their real-world counterparts.

Author Contributions: Conceptualisation, F.C., G.D.M.S. and C.C.; methodology, F.C.; software, F.C.; validation, F.C.; formal analysis, F.C.; investigation, F.C.; resources, F.C., G.D.M.S. and C.C.; data curation, F.C.; writing—original draft preparation, F.C.; writing—review and editing, F.C., G.D.M.S. and C.C.; visualisation, F.C.; supervision, G.D.M.S. and C.C.; project administration, G.D.M.S. and C.C.; funding acquisition, G.D.M.S. and C.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the SNF-ANR (Swiss and French national funds for scientific research) project No. 200021E_205720: TRACES.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data used in the scope of this study is available in open access on the portals mentioned throughout the article. The model code, data used and produced by the model, and all other codes and relevant files are available in open access in the following repository: https://gitlab.unige.ch/cas/Flann_ABM/kg-abm, accessed on 1 October 2024.

Acknowledgments: The authors wish to thank the entire TRACES committee for their valuable feedback and comments.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Gaube, V.; Remesch, A. Impact of urban planning on household's residential decisions: An agent-based simulation model for Vienna. *Environ. Model. Softw.* **2013**, *45*, 92–103. [CrossRef] [PubMed]
2. An, L.; Grimm, V.; Sullivan, A.; Turner, B.L., II; Malleson, N.; Heppenstall, A.; Vincenot, C.; Robinson, D.; Ye, X.; Liu, J.; et al. Challenges, tasks, and opportunities in modeling agent-based complex systems. *Ecol. Model.* **2021**, *457*, 109685. [CrossRef]

3. Leao, S.Z.; Pettit, C. Mapping Bicycling Patterns with an Agent-Based Model, Census and Crowdsourced Data. In *Proceedings of the Agent Based Modelling of Urban Systems*; Namazi-Rad, M.R., Padgham, L., Perez, P., Nagel, K., Bazzan, A., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2017; pp. 112–128. [CrossRef]
4. Crooks, A.; Malleson, N.; Manley, E.; Heppenstall, A. *Agent-Based Modelling and Geographical Information Systems: A Practical Primer*, 1st ed.; SAGE Publications Ltd.: Thousand Oaks, CA, USA, 2019.
5. Cantergiani, C.; Gómez Delgado, M. Urban growth simulation with AMEBA: Agent-based Model to residential occupation. In *Boletín de la Asociación de Geógrafos Españoles*; Asociacion Espanola de Geografia: Madrid, Spain, 2020. [CrossRef]
6. Maggi, E.; Vallino, E. Understanding urban mobility and the impact of public policies: The role of the agent-based models. *Res. Transp. Econ.* **2016**, *55*, 50–59. [CrossRef]
7. Ozel, B.; Petrovic, M. Green Urban Scenarios: A Framework for Digital Twin Representation and Simulation for Urban Forests and Their Impact Analysis. *Arboric. Urban For.* **2024**, *50*, 109–130. [CrossRef]
8. Bilal, S.; Zaatour, W.; Otano, Y.A.; Saha, A.; Newcomb, K.; Soo, K.; Kim, J.; Ginjala, R.; Groen, D.; Michael, E. CitySEIRCast: An Agent-Based City Digital Twin for Pandemic Analysis and Simulation. Available online: https://www.medrxiv.org/content/10.1101/2023.12.22.23300481v2 (accessed on 18 July 2024).
9. González-Méndez, M.; Olaya, C.; Fasolino, I.; Grimaldi, M.; Obregón, N. Agent-Based Modeling for Urban Development Planning based on Human Needs. Conceptual Basis and Model Formulation. *Land Use Policy* **2021**, *101*, 105110. [CrossRef]
10. Orsi, F. Centrally located yet close to nature: A prescriptive agent-based model for urban design. *Comput. Environ. Urban Syst.* **2019**, *73*, 157–170. [CrossRef]
11. Tsai, Y.; Zia, A.; Koliba, C.; Bucini, G.; Guilbert, J.; Beckage, B. An interactive land use transition agent-based model (ILUTABM): Endogenizing human-environment interactions in the Western Missisquoi Watershed. *Land Use Policy* **2015**, *49*, 161–176. [CrossRef]
12. Motieyan, H.; Mesgari, M.S. An Agent-Based Modeling approach for sustainable urban planning from land use and public transit perspectives. *Cities* **2018**, *81*, 91–100. [CrossRef]
13. Xu, T.; Gao, J.; Coco, G.; Wang, S. Urban expansion in Auckland, New Zealand: A GIS simulation via an intelligent self-adapting multiscale agent-based model. *Int. J. Geogr. Inf. Sci.* **2020**, *34*, 2136–2159. [CrossRef]
14. Hofmeister, M.; Bai, J.; Brownbridge, G.; Mosbach, S.; Lee, K.F.; Farazi, F.; Hillman, M.; Agarwal, M.; Ganguly, S.; Akroyd, J.; et al. Semantic agent framework for automated flood assessment using dynamic knowledge graphs. *Data-Centric Eng.* **2024**, *5*, e14. [CrossRef]
15. Hoch, M.; Gupta, S.; Wolkenhauer, O. Large-scale knowledge graph representations of disease processes. *Curr. Opin. Syst. Biol.* **2024**, *38*, 100517. [CrossRef]
16. An, G. Dynamic Knowledge Representation Using Agent-Based Modeling: Ontology Instantiation and Verification of Conceptual Models. In *Systems Biology*; Maly, I.V., Ed.; Humana Press: Totowa, NJ, USA, 2009; pp. 445–468. [CrossRef]
17. Chambers, F.; Di Marzo Serugendo, G.; Cruz, C. A DPSIR-Driven Agent-Based Model for Residential Choices and Mobility in an Urban Setting. *Sustainability* **2024**, *16*, 8181. [CrossRef]
18. Pavlos, T. Urban growth models and calibration methods: A case study of Athens, Greece. *Bull. Geogr. Socio-Econ. Ser.* **2022**, *55*, 107–121. [CrossRef]
19. Yao, Y.; Kshirsagar, M.; Vaidya, G.; Ducrée, J.; Ryan, C. Convergence of Blockchain, Autonomous Agents, and Knowledge Graph to Share Electronic Health Records. *Front. Blockchain* **2021**, *4*, 661238. [CrossRef]
20. Zafeiropoulos, A.; Fotopoulou, E.; Papavassiliou, S. Participatory Socio-Environmental Systems Modeling over Knowledge Graphs. In Proceedings of the 2021 IEEE Globecom Workshops (GC Wkshps), Madrid, Spain, 7–11 December 2021; pp. 1–6. [CrossRef]
21. Smeets, E.; Weterings, R.; Bosch, P.; Büchele, M.; Gee, D. *Environmental Indicators: Typology and Overview*; European Environment Agency: Copenhagen, Denmark, 1999.
22. Munafò, S.; Christie, D.; Vincent-Geslin, S.; Kaufmann, V. Typologie et évolution des Logiques de Choix Modal Chez les Actifs Motorisés Urbains, Genève, Lausanne, Switzerland, 2012. Available online: https://www.vd.ch/fileadmin/user_upload/organisation/dinf/sm/fichiers_pdf/Choix_modal_-_PrÃl'sentation_de_synthÃÍse.pdf (accessed on 10 October 2024).
23. Département des Infrastructures, République et Canton de Genève. MOBILITÉ: Des Actions Fortes Pour Respecter les Objectifs du Plan Climat Cantonal. Available online: https://www.ge.ch/document/28906/telecharger (accessed on 22 April 2024).
24. Grand Conseil de la République et Canton de Genève. Loi Modifiant la Loi Sur le Réseau des Transports Publics (LRTP) (12553). Available online: https://ge.ch/grandconseil/data/loisvotee/L12553.pdf (accessed on 4 June 2024).
25. République et Canton de Genève. SIL Genève PUBLIC-Loi Sur le Réseau des Transports Publics (LRTP) H 1 50. Available online: https://silgeneve.ch/legis/index.aspx (accessed on 4 June 2024).
26. Pays de Gex agglo. Prolongement du Tramway à Ferney-Voltaire Depuis Genève—Mobilités | Transports. Available online: https://www.paysdegexagglo.fr/ficheaction/96/8834-prolongement-du-tramway-a-ferney-voltaire-depuis-geneve.htm (accessed on 22 April 2024).
27. Wikipédia. Ligne 13 du Tramway de Genève. Available online: https://fr.wikipedia.org/w/index.php?title=Ligne_13_du_tramway_de_Gen%C3%A8ve&oldid=210179936 (accessed on 22 April 2024).
28. Wikipédia. Ligne 14 du Tramway de Genève. Available online: https://fr.wikipedia.org/w/index.php?title=Ligne_14_du_tramway_de_Gen%C3%A8ve&oldid=210870298 (accessed on 22 April 2024).

29. 20 minutes. Extension du Tram 15: Maudet Répond Aux Craintes Des Frontaliers. Available online: https://www.20min.ch/fr/story/genevefrance-voisine-extension-du-tram-15-maudet-repond-aux-craintes-des-frontaliers-103092778 (accessed on 26 April 2024).
30. Léman Bleu tv. L'extension du Tram 15 Inaugurée. Available online: https://www.lemanbleu.ch/fr/Actualites/Geneve/L-extension-du-tram-15-inauguree.html (accessed on 22 April 2024).
31. Wikipédia. Ligne 15 du Tramway de Genève. Available online: https://fr.wikipedia.org/w/index.php?title=Ligne_15_du_tramway_de_Gen%C3%A8ve&oldid=211900528 (accessed on 22 April 2024).
32. Ingerop–Uguet–Folia–Citec–T-Ingenierie . Dossier D'Enquete Publique Prealable A LA Declaration D'Utilite Publique. Available online: https://www.tram-pietonnisation.fr/wp-content/uploads/2016/09/Dossier_Tramway_Partie_1.pdf (accessed on 10 October 2024).
33. Le Dauphiné Libéré. En Images. Tram et Piétonnisation: Le Centre-Ville d'Annemasse Quasi Méconnaissable, et C'est Pas Fini! Available online: https://www.ledauphine.com/economie/2024/02/17/tram-et-pietonnisation-le-centre-ville-d-annemasse-quasi-meconnaissable-et-c-est-pas-fini (accessed on 22 April 2024).
34. Le Tram Piétonnisation-Annemasse. Info Travaux. Available online: https://www.tram-pietonnisation.fr/info-travaux/ (accessed on 22 April 2024).
35. République et Canton de Genève. Mise en Service Intégrale du Léman Express: Mesures D'Accompagnement. Available online: https://www.ge.ch/document/13455/telecharger (accessed on 22 April 2024).
36. Wikipédia. Ligne 17 du Tramway de Genève. Available online: https://fr.wikipedia.org/w/index.php?title=Ligne_17_du_tramway_de_Gen%C3%A8ve&oldid=210212545 (accessed on 22 April 2024).
37. Wikipédia. Ligne 18 du Tramway de Genève. Available online: https://fr.wikipedia.org/w/index.php?title=Ligne_18_du_tramway_de_Gen%C3%A8ve&oldid=211392161 (accessed on 22 April 2024).
38. Taillandier, P.; Gaudou, B.; Grignard, A.; Huynh, Q.N.; Marilleau, N.; Caillou, P.; Philippon, D.; Drogoul, A. Building, composing and experimenting complex spatial models with the GAMA platform. *GeoInformatica* **2019**, *23*, 299–322. [CrossRef]
39. Barat, S.; Parchure, R.; Darak, S.; Kulkarni, V.; Paranjape, A.; Gajrani, M.; Yadav, A.; Kulkarni, V. An Agent-Based Digital Twin for Exploring Localized Non-pharmaceutical Interventions to Control COVID-19 Pandemic. *Trans. Indian Natl. Acad. Eng.* **2021**, *6*, 323–353. [CrossRef]
40. Brudermann, T.; Hofer, C.; Yamagata, Y. Agent-Based Modeling—A Tool for Urban Resilience Research? In *Urban Resilience: A Transformative Approach*; Yamagata, Y., Maruyama, H., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 135–151. [CrossRef]
41. Combs, T.; Nelson, K.L.; Luke, D.; McGuire, F.H.; Cruden, G.; Henson, R.M.; Adams, D.R.; Hoagwood, K.E.; Purtle, J. Simulating the role of knowledge brokers in policy making in state agencies: An agent-based model. *Health Serv. Res.* **2022**, *57*, 122–136. [CrossRef]
42. Gerst, M.D.; Wang, P.; Roventini, A.; Fagiolo, G.; Dosi, G.; Howarth, R.B.; Borsuk, M.E. Agent-based modeling of climate policy: An introduction to the ENGAGE multi-level model framework. *Environ. Model. Softw.* **2013**, *44*, 62–75. [CrossRef]
43. Hofer, C.; Jäger, G.; Füllsack, M. Large scale simulation of CO_2 emissions caused by urban car traffic: An agent-based network approach. *J. Clean. Prod.* **2018**, *183*, 1–10. [CrossRef]
44. Kickhöfer, B.; Kern, J. Pricing local emission exposure of road traffic: An agent-based approach. *Transp. Res. Part D Transp. Environ.* **2015**, *37*, 14–28. [CrossRef]
45. Kim, D.; Batty, M. Modeling Urban Growth: An Agent Based Microeconomic Approach to Urban Dynamics and Spatial Policy Simulation. Ph.D. Thesis, University College London, London, UK, 2012
46. Maggi, E.; Vallino, E. Price-based and motivation-based policies for sustainable urban commuting: An agent-based model. *Res. Transp. Bus. Manag.* **2021**, *39*, 100588. [CrossRef]
47. Qin, C. Modeling Commute Behavior Dynamics in Response to Policy Changes: A Case Study from the COVID-19 Pandemic. Ph.D. Thesis, UC Irvine, Irvine, CA, USA, 2023.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

applied sciences

MDPI

Article

Node Importance Evaluation of Urban Rail Transit Based on Signaling System Failure: A Case Study of the Nanjing Metro

Junhong Hu *, Mingshu Yang, Yunzhu Zhen and Wenling Fu

School of Transportation Engineering, Nanjing Tech University, Nanjing 211816, China; yangmingshu@njtech.edu.cn (M.Y.); zyz21132006@163.com (Y.Z.); sqzr@njtech.edu.cn (W.F.)
* Correspondence: happyjunhong@163.com

Abstract: Assessing the importance of nodes in urban rail transit systems helps enhance their ability to respond to emergencies and improve reliability in view of the fact that most of the existing methods for evaluating the importance of rail transit nodes ignore the disturbance effect of signaling system failures and are unable to objectively identify critical stations in specific disturbance scenarios. Therefore, this paper proposed a method for evaluating the importance of urban rail transit nodes in signaling system failure scenarios. The method was based on the research background of the signaling system failure that occurs most frequently and analyzed the network failure mechanism after the occurrence of a disturbance. The node importance evaluation indices were selected from the network topology and network operation performance in two aspects. The variation coefficient–VIKOR method was employed to comprehensively assess the significance of urban rail transit stations during signaling system failures. The Nanjing Metro network was also used as an example to evaluate the importance of network stations. The results showed that under the attack method of signaling system failure, most ECC and interlocking stations experienced significantly higher network performance losses compared to the original attack method, and a few interchange stations showed smaller performance losses. The critical stations identified based on the proposed method are mainly distributed in the passenger flow backbone of the Nanjing Metro and were constructed in the early stage; of these, 85% are ECC stations or interlocking stations, which are easily neglected in daily management, in contrast to interchange stations with heavy passenger flow. The results of this study can provide an important reference for the stable operation and sustainable construction of urban rail transit.

Keywords: urban rail transit; complex networks; signaling system failure; node importance; variation coefficient–VIKOR method

1. Introduction

As an important infrastructure for urban transport, the urban rail network effectively relieves the pressure on ground transport and promotes environmental and economic sustainability due to its energy-saving, high-capacity, and non-polluting characteristics [1]. Currently, the scale of rail transport construction in China's major cities is expanding. According to the Ministry of Transportation and Communications, as of 31 December 2023, a total of 55 cities in mainland China had opened 306 urban rail transit operation lines, and the operation mileage exceeded 10,000 km, reaching 10,165.7 km [2]. Among them, 33 cities had three or more lines in operation, accounting for 60 percent of the total number of cities that had opened urban rail transit operations. This indicates that China's rail transport has entered the era of the interconnected and networked operation of stations and lines [3].

In the context of networked operation, urban rail transit is a complex system with complex lines, high equipment accuracy requirements, and a high frequency of train departures, which inevitably leads to frequent disturbances in daily operations [4]. In recent years, there has been a focus on urban rail disturbance incidents [5–7]. Disturbance

event risk factors have been categorized as human factors [8], facility and equipment factors [9], and environmental factors [10], according to the cause of the incident. It can be seen that the failure of any subsystem or component of this complex system can lead to serious consequences in terms of train delays, stoppages, and even network paralysis [11].

Train failure, signaling failure, platform door failure, and power system failure are the four most common types of operational emergencies in daily operations, accounting for 95% of all operational emergencies [12]. Among these, signaling system failures had the highest frequency of occurrence among all disturbance events in urban rail transit, accounting for 37% of the total number of disturbances, and they are the main cause of such events [13]. In addition to the high frequency of the signaling system's own failures, human causes, mechanical failures, and environmental factors can also lead to the occurrence of signaling failures, resulting in a high frequency of signaling failures. According to New York Metro data, out of 125 morning rush hours in the first half of 2021, 104 (or 80%) of the New York Metro experienced train delays due to signaling problems [14]. According to Beijing Metro data, in the five years from 2014 to 2018, the Beijing Metro accumulated 405 primary failures of various types. All of these failures caused delays ranging from a few minutes to a few hours or even led to more serious consequences. Among these, 198 signaling failures occurred, which accounted for nearly half of them [15]. It can be seen that the frequent occurrence of signaling system failures seriously affects the safe and stable operation of urban rail transit, greatly impacting urban residents' travel. Therefore, if the critical nodes in the urban rail transit network can be found and protected in advance under the disturbance scenarios of signal system failure, it not only enhances the technical safety of the rail transit network but also helps to guarantee the network operation safety and improve the service quality. This paper intends to propose a node importance assessment method under the signaling system failure scenario and apply it to the actual case of the Nanjing Metro to identify the critical stations. Furthermore, the proposed method provides a framework for assessing the significance of nodes in practical applications across various metro networks, including those in New York, London, Beijing, Shanghai, and other extensive rail systems.

The rest of this paper proceeds as follows: Section 2 gives an overview of the research on node importance evaluation and network failure mechanisms; Section 3 provides an overview of the construction of an urban rail transit network model and defines the network failure mechanisms in signaling system failure scenario; in Section 4, an evaluation method of node importance is constructed, and the selection of evaluation indices and the comprehensive evaluation method of node importance is described in detail; Section 5 applies the critical station identification method proposed in this paper to the Nanjing Metro network and analyzes the results; and finally, conclusions and future work are provided in Section 6.

2. Literature Review

Node importance evaluation has been a hot research topic in the field of urban rail transport. In recent years, scholars have conducted some case studies on the evaluation of the importance of nodes in real rail networks. Amideo et al. [16] evaluated the importance of urban rail nodes using the London Underground network as a practical example. Kilic et al. [17] conducted a comprehensive evaluation of the rail lines of the Republic of Turkey State Railways (TCDD). Peng et al. [18] used the metro network in Suzhou, China, to evaluate the node importance of the metro network.

In terms of evaluation indices, scholars initially selected node degree [19], facility service level [20], and other factors to evaluate rail transit stations from a local perspective. Subsequently, some scholars adopted the complex network theory from a global perspective, using node degree centrality, betweenness centrality, closeness centrality [21,22], and other indicators to evaluate the importance of nodes in static urban rail transit networks [23,24]. The evaluation method can only reflect the performance of the network when it is in a stationary state (that is, under normal operation), and cannot reflect the change in the

network state of urban rail transit in actual operation. For example, the degree can only reflect the distance between nodes, and the degree of centrality can only reflect the existence of connecting edges between nodes.

However, the urban rail transit network is dynamically changing in actual operation, and considering the actual operation of rail transit, some studies have introduced the network interdiction model [25], which simulates the network failure by removing nodes and edges in the network. The network interdiction model was first introduced for military purposes, intending to identify and determine the most critical set of arcs in a network [26,27], and has subsequently been extended to the domain of complex networks, such as electric power networks [28], wireless sensor networks [29], and rail transit networks [25]. Combined with the actual operation of rail transit, the existing research constructed a node importance evaluation system based on the dynamic perspective, selecting indicators such as average network efficiency [30], maximum connectivity subgraph change [31], network efficiency accessibility [32,33], and unsatisfied passenger demand [34,35] from the aspects of network topology and network operation performance. Then, the network interdiction model was used to carry out station failure simulations in urban rail transit networks to carry out network node importance evaluations and key station identification.

In terms of network failure mechanisms, most of the studies only remove the failed station itself and its connected lines from the network when simulating station failure [25,36], while a few studies improve the failure mechanism by taking into account the actual characteristics of the rail network and disruption scenarios. Chen et al. [37] considered that urban rail transit operates on specific routes, used spatial syntax to analyze the line accessibility of rail transit to improve the topological network, and proposed the node importance evaluation method. Hu et al. [38] incorporated the adjustment of train operation interchanges after station failure to evaluate the importance of rail transit stations in the scenario of bidirectional operation disruption. Zhang et al. [39] proposed a weighted multi-layer network node importance evaluation method that considers node connectivity, diverse transportation modes, and service capabilities. Zheng et al. [40], combined with the unexpected situations encountered in the actual operation of public transport, considered two types of attacks against the bus–subway composite network nodes for node importance evaluation research.

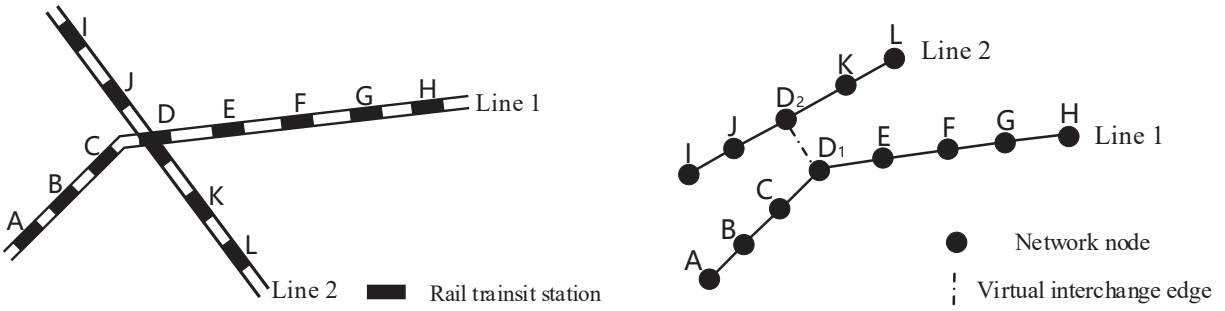
However, the urban rail transit system is a complex system containing many subsystems, in which normal operation is built on the basis of mutual cooperation and the stable operation of facilities and equipment among the subsystems. In particular, the signaling system, as the key to the signal transmission of the rail transport network and the component that ensures that the trains run according to the plan, is an important guarantee for the safety of rail transport [41]. If the critical stations where the facilities and equipment are located are not identified and maintained in advance, it will lead to a series of chain reactions once it fails. From the summary of the above literature, there are few studies that consider the failure mechanism of signaling system failures in the evaluation of the importance of urban rail transit stations.

To address the aforementioned issues, this paper fully considers the structural characteristics of urban rail transit networks and proposes a method to evaluate the importance of urban rail transit nodes under the signaling system failure disturbance scenario. Firstly, based on complex network theory, the improved Space L method is used to construct the urban rail transit network. Then, the signaling system failure scenario is taken as the research background to analyze the failure mechanism of the urban rail transit network after the occurrence of the perturbation. Finally, according to the characteristics of urban rail transit, the indicators are selected from two aspects of network topology and network operation performance, and the evaluation method of the importance of urban rail transit network nodes based on the coefficient of variation–VIKOR method iss proposed. The greatest innovation of this approach is that it considers the structural characteristics of the rail network, as well as the network failure mechanisms during disturbance events. In this study, the Nanjing Metro network was selected as a study case.

3. Network Failure Mechanism in Signaling System Failure Scenario

3.1. Urban Rail Transit Network Model

Based on the complex network theory [42], the Space L method, which can better reproduce the actual topology of the network, is used to construct the urban rail transit network model. With the Space L method, the rail transit stations are abstracted as nodes, and the lines between adjacent stations are abstracted as connecting edges for modeling [43]. However, in actual operation, different rail transit lines are relatively independent, so this paper assumed that the failure of one line in the same interchange station does not affect the operation of another line and proposed the virtual interchange edge method [44] to construct the rail transit network model, which splits the interchange stations into multiple virtual nodes connected with each other by virtual edges. The specific network model construction method is shown in Figure 1.



(**a**) Wiring diagram of the actual urban rail transit network (**b**) Space L method with introduction of virtual interchange edge

Figure 1. Urban rail transit network modeling method.

3.2. Network Failure Mechanism

Established studies mostly focused on low-frequency and high-loss types of man-made disturbances and natural disturbances, while less consideration was given to high-frequency low-loss types of technical disturbances. According to China's urban rail transit data statistics, signaling system failure accounts for 42.53% of technical disturbance events in urban rail transit [13]. On 27 September 2011, after a signal failure at Xintiandi Station on Line 10 of the Shanghai Metro, the telephone occlusion method was used, resulting in a rear-end collision of the train and injuring 271 passengers [45]. All these data indicate that although the disturbance impact and damage level of signaling system failures can be easily predicted, they occur with high frequency, and when they occur, they must also be technically overhauled before resuming operation. Therefore, the problem of signaling systems affecting the stable operation of rail transit needs to be solved urgently. In this paper, the technical disturbance scenario of signaling system failure was used as the background to analyze the network failure mechanism of urban rail transit after the failure, and the evaluation of station importance is carried out on this basis.

Interlocking system for urban rail transport refers to the mutual constraints established by technical means between signaling machines, turnouts, and approaches to ensure the safety of running and shunting operations in railway stations [46]. The equipment that realizes this relationship is called interlocking equipment, and according to the interlocking equipment, rail transit stations can be divided into interlocking stations, element control computer (ECC) stations, and non-centralized stations. Among these, the ECC uses the component control computer to extend the operation capability and control range of its signaling system. Its basic function is to send and issue control commands to external execution devices, receive dynamic information from these devices, and feed it back into the interlocking computer to complete the feedback process of information [47]. That is, it outputs commands and collects information to all non-centralized stations within the ECC boundary and sends the dynamic information received back to the interlocking station. The role of the interlocking station is to transmit and receive information from all ECC

stations within the interlocking boundary. The schematic diagram of an urban rail transit signaling interlocking system is shown in Figure 2.

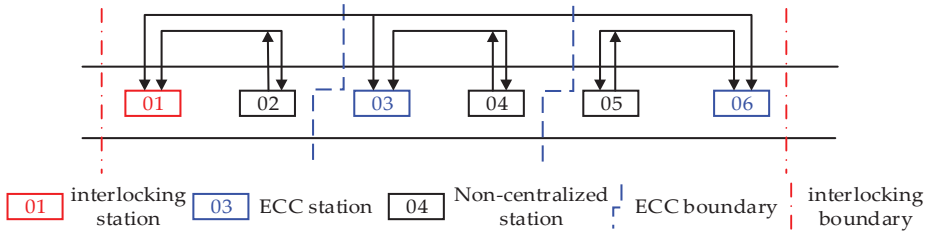


Figure 2. Schematic diagram of urban rail transit signaling interlocking system.

Based on the interlocking relationship of urban rail transit signaling systems, the network failure mechanism is defined. When signaling system failure occurs in urban rail transit, there are three scenarios. If the ECC station fails, it will cause the station and line within the control range of the ECC station (within the ECC boundary) to lose control of the outdoor executive equipment and stop receiving dynamic information, making the station and line within the ECC boundary fail. If the interlocking station fails, the station and line within the control range of the interlocking station (within the interlocking boundary) will not be able to receive and feed back the information, which means that the stations and lines within the entire interlocking boundary fail. If the non-centralized station fails, the station will not be able to communicate with the ECC station, which means that the station and its connected line fail. Based on the above analyses, we know that the signaling failure incident at Xintiandi Station of Shanghai Metro Line 10 in 2011 had a huge impact on network operation and passenger travel.

Figure 3 provides an example of the network model before and after station attacks. Figure 3a,b depict the actual wiring diagram and the abstract network model, respectively. Figure 3c,d simulate two types of attacks resulting from signaling system failures. It can be seen that the original attack method only considers the failure station itself, ignoring the structural characteristics of the rail transit itself and the failure mechanism after a disturbance, so the rail transit network model after the attack is not compatible with reality.

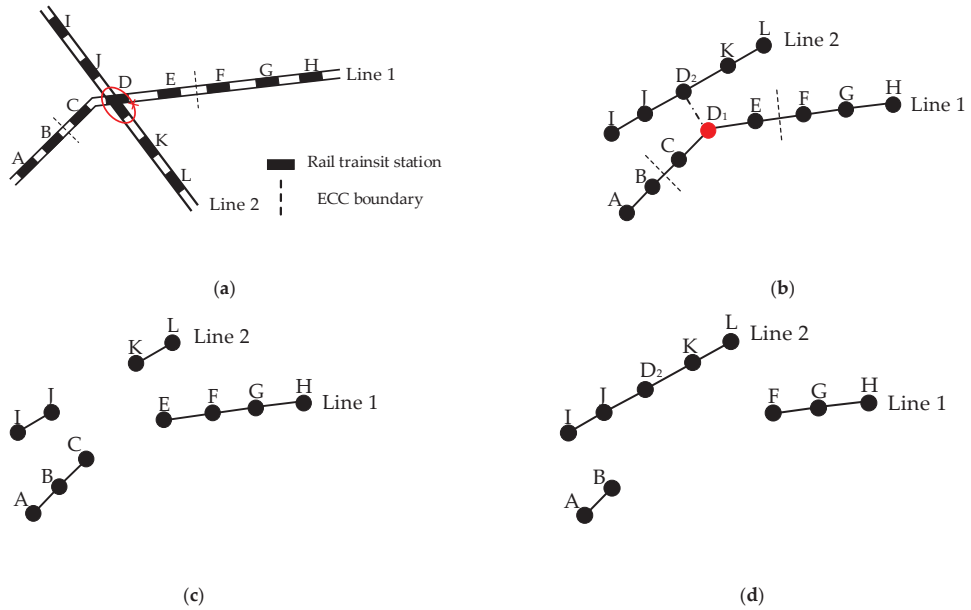


Figure 3. Network model before and after the attack on the stations. (**a**) Actual wiring diagram (assuming the signaling system failure at ECC station D on line 1); (**b**) Modeling of the rail transit network before the attack; (**c**) Modeling the rail transit network under the original attack approach; (**d**) Modeling rail transit networks under the attack approach of this paper.

4. Node Importance Evaluation Model

4.1. Indices for Evaluating the Importance of Urban Rail Transport Nodes

Compared to static evaluation, dynamic metrics are able to capture the change in network performance after site removal, thus providing a more accurate measure of the impact of site failure on the network and evaluating site importance. Therefore, in this paper, we mainly select two dynamic indicators to measure the impact of site failure: network topology and network operational performance.

4.1.1. Subsubsection

The network efficiency loss and maximum connected sub-network loss after the failure of urban rail transit stations are selected as dynamic evaluation indices to reflect the impact of station failure on network topology performance.

(1) Network efficiency loss

The network efficiency of urban rail transit, as a convenient and efficient mode of public transport, quantifies the efficiency of the network on a global scale. After the failure of an urban rail station, the shortest paths between certain starting and finishing points in the network changed, resulting in efficiency losses for the network. The loss of network efficiency is shown in Formulas (1) and (2).

$$E = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{ij}} \tag{1}$$

$$\Delta E_a = E_0 - E_a \tag{2}$$

E is the network efficiency, N is the total number of stations in the network, d_{ij} is the shortest path length between stations i and j. Then, E_0 is the initial network efficiency, E_a is the efficiency of the network after the failure of station a, and ΔE_a is the loss of network efficiency before and after the failure of station a.

(2) Maximum connected sub-network loss

After the failure of urban rail transit network stations, the network becomes divided into different connected sub-networks. The sub-network with the highest number of connections is referred to as the maximum connectivity sub-network. The calculation of the maximum connected sub-network loss can measure the impact of the station failure on network connectivity. The loss of the maximum connected sub-network is represented by Formula (3).

$$\Delta S_a = 1 - \frac{N_a}{N} = \frac{N - N_a}{N} \tag{3}$$

N_a is the number of stations in the maximum connectivity sub-map of urban rail transit after the failure of station a, and ΔS_a is the maximum connected sub-network loss before and after the failure of station a.

4.1.2. Based on Network Operational Performance

Considering the importance of passenger traveling services in the operation of rail transit networks, network efficiency accessibility loss and unsatisfied passenger demand are selected as dynamic evaluation indices to reflect the impact of station failure on network operational performance.

(1) Network efficiency accessibility loss

Network efficiency accessibility is based on the network efficiency indicators, adding the consideration of the passenger flow between the stations in the network, which can reflect the difficulty of different passengers in the urban rail transit network to reach the

destination station from the initial station [33], and network efficiency accessibility loss is shown in Formulas (4) and (5).

$$A = \frac{1}{f}\sum_{i}^{n}\sum_{j}^{n-1}\frac{f_{ij}}{t_{ij}} \tag{4}$$

$$\Delta A_a = A_0 - A_a \tag{5}$$

A is the network efficiency accessibility, f_{ij} is the flow of passengers from station i to j, f is the total number of passengers in the network, t_{ij} is the time it takes a passenger to travel from station i to j, A_0 is the initial efficiency accessibility of the network, A_a is the network efficiency accessibility after the failure of station a, and ΔA_a is the loss of network efficiency accessibility before and after the failure of station a.

(2) Unsatisfied passenger demand

The disruption of some stations in the rail transit network splits the network into several disconnected sub-networks, resulting in the absence of connected paths between the origin and destination points of some passengers' trips, and the passengers are unable to complete their trips by the rail transit, which is the unsatisfied passenger demand, and the expression is shown in Formula (6).

$$U = \frac{f_a}{f} \tag{6}$$

f_a indicates passengers who are unable to complete their journeys using urban rail transit after the failure of station a, and U indicates unmet passenger demand after the failure of station a.

4.2. Comprehensive Evaluation Based on the Variation Coefficient–VIKOR Method

Multiple criteria decision-making (MCDM) is an important technique that presents a systematic approach to helping decision-makers in this field [48]. The MCDM method includes AHP, TOPSIS, VIKOR, and PROMETHEE, among others. Among these, the basic idea of the VIKOR method is to normalize the original data, construct the initial matrix, select the optimal and the worst solution in the whole solution set, and determine the comprehensive ranking results according to the distance between the optimal and the worst solution between different values, that is, the criterion that the optimal solution is the closest, and the worst solution is the furthest [49]. The method is often a compromise between the attributes so that the feasible solution can ensure the maximum benefit of the group but also take into account the individual loss, which can effectively deal with the problem of urban rail transit station importance based on the evaluation of a number of indicators and make full use of the original data in the calculation process. Therefore, the VIKOR method is selected in this paper for the comprehensive evaluation of station importance.

However, the VIKOR method defaults to equal weighting of all evaluation indicators, which is not consistent with the actual situation. This paper adopted the coefficient of variation method, which is an objective weighting method, to enhance the VIKOR comprehensive evaluation method. The improved method evaluates the degree of the data dispersion of indicators as an assignment criterion, assigning higher weights to indicators with greater variation dispersion [38]. The specific evaluation process is as follows.

4.2.1. Determination of Indicator Weights

Assuming that there are evaluation objects and evaluation indicators, the coefficient of variation method is used to determine the weights of the indicators, which are calculated as shown in Formulas (7) and (8).

$$v_k = \frac{\sigma_k}{\overline{x_k}} = \frac{1}{\overline{x_k}}\sqrt{\frac{1}{m}\sum_{i=1}^{m}(x_{ik} - \overline{x_k})^2} \tag{7}$$

$$w_k = \frac{v_k}{\sum_{k=1}^{n} v_k} \tag{8}$$

x_{ik} is the value of the ith evaluation object for the kth indicator (i = 1,2,$\cdots$,m; k = 1,2,$\cdots$,n) , $\overline{x_k}$, σ_k is the mean and mean square of the jth indicator for all stations, and v_k is the coefficient of variation of the kth indicator.

4.2.2. Comprehensive Evaluation of the VIKOR Method

(1) Normalizing the raw data

Normalizing the raw data yields matrix X, as shown in Formula (9).

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \tag{9}$$

(2) Determining positive and negative ideal solutions for each indicator

Positive ideal solutions x_k^+ and negative ideal solutions x_k^- are shown in Formulas (10) and (11).

$$x_k^+ = \{\max x_{1k}, \max x_{2k}, \cdots, \max x_{nk}\} \tag{10}$$

$$x_k^- = \{\min x_{1k}, \min x_{2k}, \cdots, \min x_{nk}\} \tag{11}$$

(3) Determining group utility values S_i and individual regret values for each evaluator R_i

When the positive ideal solution is used as a reference, the group utility value S_i^+ and individual regret value R_i^+ of each evaluation object are calculated, as shown in Formulas (12) and (13).

$$S_i^+ = \sum_{i=1}^{m} \left(w_k \frac{x_k^+ - x_{ik}}{x_k^+ - x_k^-} \right) \tag{12}$$

$$R_i^+ = \max \left(w_k \frac{x_k^+ - x_{ik}}{x_k^+ - x_k^-} \right) \tag{13}$$

When the negative ideal solution is used as a reference, the group utility value S_i^- and individual regret value R_i^- of each evaluation object are calculated, as shown in Formulas (14) and (15).

$$S_i^- = \sum_{i=1}^{m} \left(w_k \frac{x_{ik} - x_k^-}{x_k^+ - x_k^-} \right) \tag{14}$$

$$R_i^- = \min \left(w_k \frac{x_{ik} - x_k^-}{x_k^+ - x_k^-} \right) \tag{15}$$

In summary, the group utility value S_i and individual regret value R_i of each evaluation object are calculated as shown in Formulas (16) and (17).

$$S_i = \frac{S_i^+}{S_i^-} \tag{16}$$

$$R_i = \frac{R_i^+}{R_i^-} \tag{17}$$

(4) Calculating the discounted assessed value of the subject of the evaluation Q_i

$$Q_i = v\frac{S_i - S_{i\min}}{S_{i\max} - S_{i\min}} + (1-v)\frac{R_i - R_{i\min}}{R_{i\max} - R_{i\min}} \tag{18}$$

where v is the coefficient of the decision-making mechanism, $0.5 < v \leq 1$ indicates that the evaluator prefers group utility to individual regret, and $0 \leq v < 0.5$ indicates that the evaluator prefers individual regret. In order to maximize group utility and minimize individual regret at the same time, this paper chooses $v = 0.5$ [49]. The value of the compromise evaluation is used as the evaluation value of the importance of the stations of the urban rail transit network in this paper, and the smaller the value of the compromise evaluation is, the higher the importance of that station is indicated.

5. Case Study

5.1. Nanjing Metro Network

In this paper, we select the Nanjing Metro network as an example and apply the method proposed in this paper to evaluate the station's importance in a signaling system failure scenario.

As of September 2023, Nanjing has opened 12 metro lines for operation, with a total length of about 449 km and containing a total of 208 stations (with double counting of interchange stations) and 217 line intervals, and has entered the stage of networked operation. The Nanjing Metro network is shown in Figure 4.

Figure 4. Nanjing Metro network (September 2023).

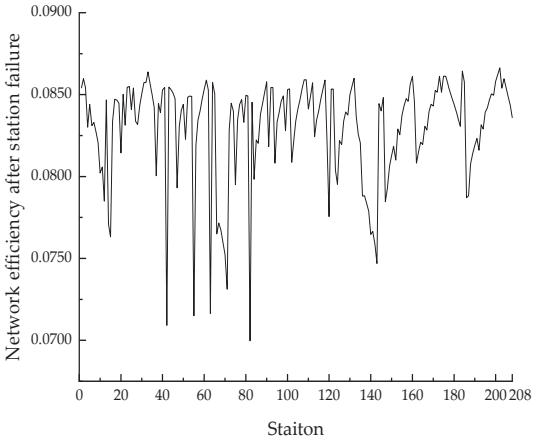
This paper was based on the Nanjing Metro line wiring diagram, and the Nanjing Metro passenger flow data extracted from the AFC system was analyzed. We obtained information on the signaling system facilities and equipment of each line from the metro line wiring diagram and found that there are 60 interlocking stations, 37 ECC stations, and 111 non-centralized stations in the Nanjing Metro. The AFC data used are the data from the set counting method, including the average daily OD between stations and the passenger flow in and out of the stations based on the direct counting of the swipe card data from

the entrance and exit gates of the stations and the passenger flow of the interchanging passengers calculated using the clearing model of the metro company.

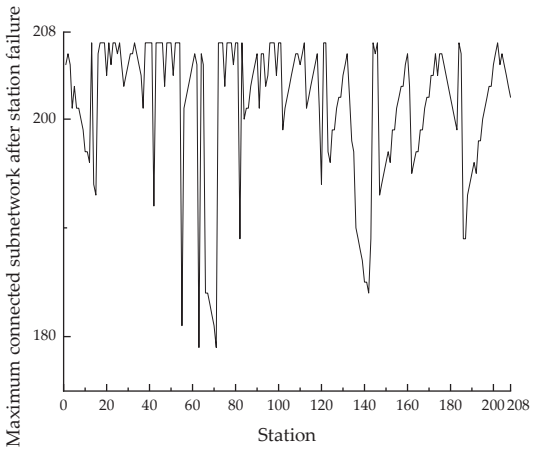
5.2. Evaluation of the Importance of Urban Rail Network Nodes

5.2.1. Calculation of Evaluation Indicator Values

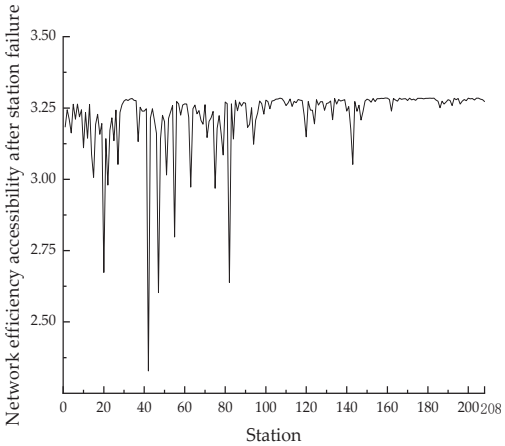
Based on the Space L method, with the introduction of a virtual interchange edge, the urban rail transit network model has been constructed. Figure 5 presents the impact on the network after the station failure using the attack method in this paper in terms of the network efficiency, maximum connected sub-network, network efficiency accessibility, and level of unsatisfied passenger demand after the station failure. We found that there is a big difference in the impact of the different station failures on the network topology and operation performance. This is also in line with the actual operation of urban rail networks. For example, the 47th station, Xinjiekou, is located in the core business area of Nanjing and is a large interchange station serving 192,218 passengers each day. When the Xinjiekou station (ECC station) of Line 2 fails, there will be a loss of network efficiency in terms of network topology of 0.0078 and a loss of maximum connectivity subgraph of 0.0240. In terms of passenger traveling service, 373,385 passengers will not be able to satisfy their traveling needs via urban rail, with a loss of network efficiency accessibility of 0.6823 and a loss of unmet passenger demand of 0.1667.



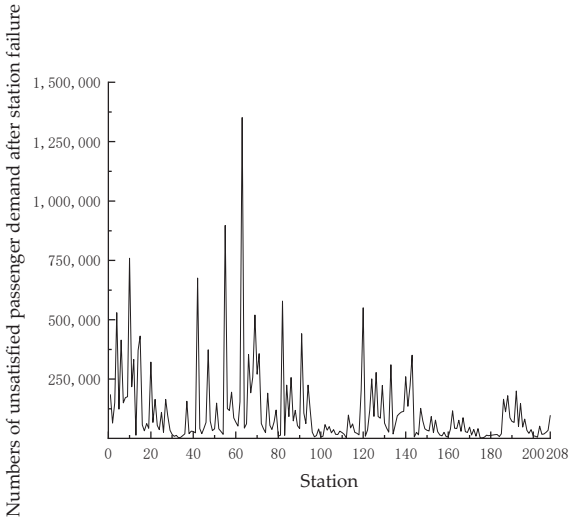
(**a**) Network efficiency after station failure



(**b**) Maximum connected sub-network after station failure



(**c**) Network efficiency accessibility after station failure



(**d**) Unsatisfied passenger demand after station failure

Figure 5. Impact of station failure on the network.

Table 1 lists 10 representative station indicators, with values calculated according to Formulas (1)–(6). These values represent the indicators of each station in the urban rail transit network.

Table 1. Example of values for evaluation indicators.

Station Type	Station Name	Network Efficiency Loss		Maximum Connected Sub-Network Loss		Network Efficiency Accessibility Loss		Unsatisfied Passenger Demand	
		The Original Way	This Paper	The Original Way	This Paper	The Original Way	This Paper	The Original Way	This Paper
Interlocking station	China Pharmaceutical University Station (Line 1)	0.0005	0.0017	0.0005	0.0144	0.0372	0.1017	0.0151	0.0825
Interlocking station	Tianlongsi Station (Line 1)	0.0022	0.0108	0.0048	0.0721	0.0901	0.2786	0.0269	0.1927
Interlocking station	Gulou Station (Line 1)	0.0053	0.0139	0.0240	0.0144	0.2145	0.3057	0.0851	0.0733
Interlocking station	Nanjing South Station (Line S3)	0.0232	0.0124	0.2163	0.0913	0.4986	0.2333	0.3931	0.1566
ECC station	Xinjiekou Station (Line 2)	0.0069	0.0078	0.0096	0.0240	0.6643	0.6823	0.1414	0.1667
ECC station	Xiaoshi Station (Line 3)	0.0034	0.0139	0.0048	0.1394	0.0365	0.1387	0.0108	0.0595
ECC station	Ruanjiandadao Station (Line 1)	0.0022	0.0100	0.0048	0.0673	0.0376	0.1981	0.0111	0.1658
Non-centralized stations	Zhujianglu Station (Line 1)	0.0021	0.0021	0.0048	0.0048	0.1426	0.1426	0.0303	0.0303
Non-centralized stations	Xuanwumen Station (Line 1)	0.0016	0.0016	0.0048	0.0048	0.1133	0.1133	0.0245	0.0245
Non-centralized stations	Yuantong Station (Line2)	0.0085	0.0032	0.0577	0.0048	0.0896	0.0464	0.0670	0.0139

In Table 1, it can be seen that in the attack method based on a signaling system failure scenario, the values of the indicators for most of the non-centralized stations did not change. Most of the ECC and interlocking stations, such as China Pharmaceutical University Station and Xinjiekou Station, have values that are significantly larger than those of the original attack method, which indicates that the original attack method ignores the network failure mechanism during signaling system failure and weakens the negative impact of some station failures on network performance. However, there are also a few interchange stations, such as Gulou Station, Nanjing South Station, and Yuantong Station, where the values of the metrics are smaller than those of the original attack method. This is because under the original attack method, the characteristics of the urban rail network in which each line operates independently of each other are ignored, and when the interchange station fails, the station and its connected lines are all removed from the network, which overestimates the impact of its failure on the network.

5.2.2. Calculation of Evaluation Indicator Weights

Combined with the value of each evaluation index, according to Formulas (7) and (8), the coefficient of variation method is used to calculate the weight of each index, as shown in Table 2.

Table 2. Weights of evaluation indicators.

Indicator type	Indicators	Weights
Network topology	Network efficiency loss	0.1499
	Maximum connected sub-network loss	0.1897
Network operational performance	Network efficiency accessibility loss	0.3727
	Unsatisfied passenger demand	0.2877

5.2.3. Results and Analysis of Node Importance Evaluation

Based on the actual fault failure mechanism of the urban rail transit signaling system, the group utility value and individual regret value of each rail transit station are obtained according to Formulas (9)–(17), and the compromise evaluation value of each station is calculated by combining Formula (18). The compromise evaluation value is ranked from low to high to obtain the results of the ranking of the importance of each station of the Nanjing Metro network. The top 20 stations in terms of node importance are listed as critical stations, as shown in Table 3.

Table 3. Critical stations identified based on signaling system fault failure mechanism.

Station Type	Station Name (Line)	Compromise Evaluation Value	Station Ranking (The Original Attack Method)	Station Type	Station Name (Line)	Compromise Evaluation Value	Station Ranking (The Original Attack Method)
Interlocking station	Jiqingmendajie Station (Line 2)	0.0000	1(32)	Interlocking station	Gulou Station (Line 1)	0.6097	11(4)
Interlocking station	Daminglu Station (Line 3)	0.1196	2(117)	ECC station	Ruanjiandadao Station (Line 1)	0.6147	12(81)
Interlocking station	Maqun Station (Line 2)	0.1284	3(9)	Interlocking station	Hedingqiao Station (Line 1)	0.6178	13(26)
Interlocking station	Linchang Station (Line 3)	0.2505	4(154)	ECC station	Muxuyuan Station (Line 2)	0.6417	14(41)
ECC station	Xinjiekou Station (Line 2)	0.3531	5(2)	ECC station	Shangyuanmen Station (Line 3)	0.6505	15(16)
Interlocking station	Xinjiekou Station (Line 1)	0.4197	6(2)	Interlocking station	Andemen Station (Line 10)	0.6605	16(49)
Interlocking station	Tianlongsi Station (Line 1)	0.5080	7(24)	Non-centralized stations	Nanjing South Station (Line S1)	0.6820	17(1)
Interlocking station	Nanjing South Station (Line S3)	0.5421	8(1)	Interlocking station	Maigaoqiao Station (Line 1)	0.6915	18(27)
ECC station	Fuqiao Station (Line 3)	0.5632	9(56)	Non-centralized stations	Nanjing South Station (Line 1)	0.7009	19(1)
ECC station	Xiaoshi Station (Line 3)	0.5919	10(73)	Non-centralized stations	Liuzhoudonglu Station (Line 3)	0.7116	20(7)

As shown in Table 3, there are 11 interlocking stations and 6 ECC stations in the top 20 stations in terms of node importance, with 85% of the total number of critical stations. Each interlocking sub-district manages multiple stations, which will indirectly lead to the disability of the stations in the whole interlocking sub-district to operate normally once its signaling system fails. For example, the Jiqingmen Street interlocking zone manages 16 stations in the Yuzui–Daxinggong zone, and the failure of this interlocking station will prevent 675,576 passengers across the region from completing their journeys on the metro, so its importance is very high. In reality, this conclusion has also been verified; for example, on 6 November 2023 at 6:54 A.M. during the morning high peak, the Nanjing Metro interlocking station, Andemen Station, suffered a sudden signaling equipment failure, which resulted in a temporary stop of the train, affecting a large number of passengers' commuting trips [50]. Some non-interlocked stations, such as the Nanjing South Station of Lines 1, S1, and S3, have a greater impact on the Nanjing Metro network, and even the transport network, which is due to their transport hub properties, making their importance relatively high.

Combined with Table 2, it can be seen that there are large differences in the identification results of critical stations based on the two attack methods. On the one hand, it is because the original attack method ignores the relatively independent characteristics of rail transit line operation, which overestimates the importance of some interchange stations in the network. On the other hand, it is because the original attack method does not consider the network failure mechanism after the disturbance occurs, that is, the failure of the interlocking stations will lead to the disruption of the operation of other stations in their interlocking interval, so the importance of some interlocking stations is underestimated.

Figure 6 shows the distribution of the top 20 critical stations in the network of the Nanjing Metro before and after the improvement of the attack method. From Figure 6a, it can be seen that the critical stations identified based on the original attack method are mainly distributed in the transfer stations in the central area of the network and some

non-transfer stations with heavy passenger traffic. This is because the original attack method only considers a single station, and the interchange stations in the central area of the network have high traffic volume and contribute more to the network performance; thus, the assessed stations have a high degree of importance. Figure 6b shows the critical stations identified based on the failure mechanism of the signaling system; most of the critical stations are the interlocking stations of the signaling system, such as Jiqingmendajie Station and Linchang Station. The failure of such stations in the original attack method does not have a great impact on the network directly, but because the interlocking zone under their management contains other important stations, the failure of such stations has a wider impact. Most of the critical stations are located on Lines 1, 2, and 3, which are the backbone of the Nanjing Metro passenger flow, carrying most of the interchange traffic and relying on the interchange function of the metro network in the city center. Once the critical stations fail, they will form an isolated line, which will have a serious impact on the traveling of the residents along the line.

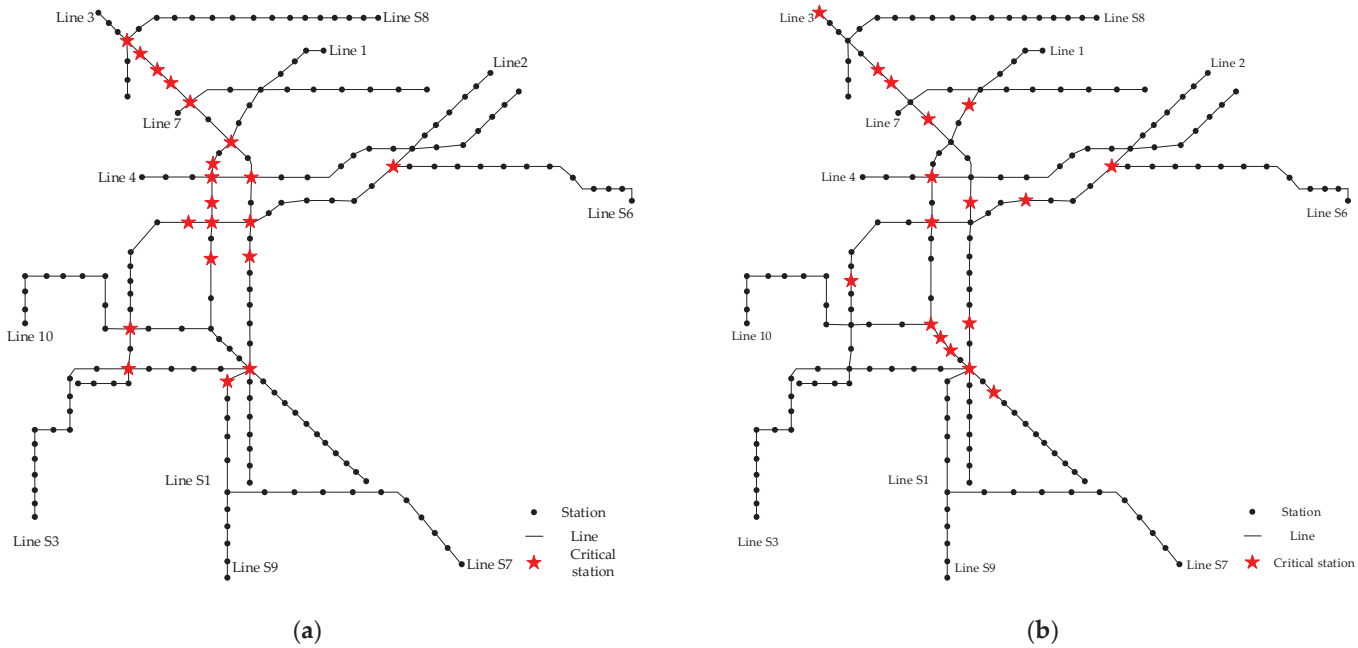


Figure 6. Distribution of critical stations. (**a**) Distribution of critical stations identified based on the original attack approach; (**b**) Distribution of critical stations identified based on the attack approach in this paper.

It can be seen that the method can not only more realistically describe the network's state after signaling a system failure but also effectively identify the critical stations that are easily ignored in other studies, as well as daily management to provide a reference for the identification and protection of critical stations in urban rail transit networks. Considering the actual structural characteristics of urban rail transit, it is necessary to determine the failure mechanism according to the signaling system in the urban rail transit network and to evaluate the critical stations of the rail transit network under the failure of the signaling system further.

6. Conclusions

This paper uses the disturbance scenario of signaling system failure as the research background and analyzes the impact on the rail transit network after the disturbance. Firstly, the Space L method with the introduction of a virtual interchange edge is used to construct the urban rail transit network. Then, indices of two aspects, network topology and network operation performance, are selected, and the comprehensive evaluation method of the variation coefficient–VIKOR method is adopted to propose the evaluation method

of the importance of urban rail transit nodes in the signaling system failure scenario. The main conclusions from the case study are as follows.

(1) This paper considered the structural characteristics and actual operation characteristics of urban rail transit, including the mutual independence of different line operations in the construction of rail transit networks and the failure mechanism of signaling system failure in the evaluation of node importance. This method can more comprehensively and objectively determine the value of the evaluation indices of each metro station and correct the impact of station failure on the network in technical disturbance scenarios.

(2) The variation coefficient–VIKOR method is used to propose a comprehensive evaluation method of site importance, incorporating passenger flow as an important factor into the dynamic index of network operation performance and combining it with the dynamic index of network topology. This made the node importance evaluation results more objective and practical.

(3) The method in this paper identified that 85% of the top 20 critical stations in terms of node importance are interlocking stations, and most of them are distributed in Nanjing Metro passenger flow backbone Lines 1, 2, and 3. Compared with the original attack method of identifying critical stations, the method in this paper can identify critical stations that are easily ignored by daily management, which are not necessarily the stations with the heaviest passenger flows or interchange stations in the metro network. The focused maintenance of these stations can minimize the loss of the network caused by the failure of the signaling system and guarantee the stable operation of the urban rail transit network.

(4) On the basis of identifying critical stations using the method proposed in this paper, for the management and maintenance of the critical stations of urban rail transport, the management department should formulate a perfect emergency response plan to deal with emergencies occurring at the identified critical stations. For example, interlocking stations and ECC stations, such as Jiqingmen Street Station, Xinjiekou Station, Nanjing South Station, etc., should strengthen the emergency training of the staff and the management of the equipment to comprehensively improve the emergency response capability of the critical stations of the rail transport and even the rail transport network. Only in this way can the occurrence of disturbance events, such as signaling failures caused by various factors, be solved.

In summary, it is recommended that the study on the evaluation of the importance of urban rail transit nodes be carried out in a specific disturbance event scenario. Considering the structural characteristics and operational characteristics of the rail transit itself, using the coefficient of variation–VIKOR method to evaluate the results is more objective and practical.

The research in this paper can help the metro operation and management departments find the critical stations in a signaling system failure and protect them in advance, which can effectively reduce the frequency of failure and its negative impacts and further ensure the stable operation of the rail transit and the safe travel of passengers. In addition, the method proposed in this paper has wide applicability and transferability. It can be widely applied to other types of large rail transit networks such as those in New York and London by evaluating node importance during signaling system failures, according to their network topology and network operation performance characteristics. It is also possible to adapt the comprehensive evaluation method proposed in this paper to evaluate node importance under other disturbance event scenarios. Failure mechanisms are identified based on fault types, and thus, critical stations are identified to provide special protection for urban rail transit networks against various types of disturbance events. This helps to improve the operational management and service level of the metro.

This paper has some implications and limitations for future studies. In terms of disturbance event context, this paper only considered the signaling system failure scenario with the highest frequency of failures and analyzed the network failure mechanism for this

type of failure, while other disturbance scenarios were not considered. In future studies, the identification results can be compared with the results of node importance evaluation under other scenarios when possible. In terms of the evaluation method, we considered both network topology and network operational performance indicators to measure the network performance but did not conduct a sensitivity analysis for the change in network indicator weights to further illustrate the robustness of the research results. In terms of the research object, this paper only considered the passenger flow within the urban rail transit network, whereas urban rail transit, as part of the urban public transport network, can be combined with the surface bus network in the future to consider the passenger flow in the dynamic integrated network. These aspects can be further studied in the future to better promote the development of urban rail transit networks in the direction of safety, stability, and reliability.

Author Contributions: Project administration, J.H.; conceptualization, J.H. and M.Y.; investigation, M.Y. and Y.Z.; analysis, J.H., M.Y. and W.F; visualization J.H., M.Y., Y.Z. and W.F.; writing—original draft, J.H. and M.Y.; writing—review and editing, J.H., M.Y., Y.Z. and W.F. All authors have read and agreed to the published version of the manuscript.

Funding: The research and publication of this article were funded by the National Natural Science Fund (Grant no. 51878349) and the Postgraduate Research and Practice Innovation Program of Jiangsu Province, China (Grant no. SJCX24_0495).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the urban rail traffic data used in this paper are from the actual data of the Nanjing Metro in Jiangsu Province, further inquiries can be directed to the corresponding author. The authors guarantee the authenticity and reliability of the data used in the validation section.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

References

1. Liu, X.; Xia, H. Networking and sustainable development of urban spatial planning: Influence of rail transit. *Sustain. Cities Soc.* **2023**, *99*, 104865. [CrossRef]
2. Ministry of Transport of the People's Republic of China. Quick Facts on Urban Rail Transit Operational Data 2023. Available online: https://mp.weixin.qq.com/s?__biz=MzI3MDQwMDQ5NQ==&mid=2247593667&idx=1&sn=5a5fd0f406b50ad9307620d7e3964d35&scene=0 (accessed on 25 July 2024).
3. Meng, Y.; Zhao, X.; Liu, J.; Qi, Q. Dynamic Influence Analysis of the Important Station Evolution on the Resilience of Complex Metro Network. *Sustainability* **2023**, *15*, 9309. [CrossRef]
4. Li, M.; Wang, Y.; Jia, L.; Cui, Y. Risk propagation analysis of urban rail transit based on network model. *Alex. Eng. J.* **2020**, *59*, 1319–1331. [CrossRef]
5. Zhang, X.; Deng, Y.; Li, Q.; Skitmore, M.; Zhou, Z. An incident database for improving metro safety: The case of shanghai. *Saf. Sci.* **2016**, *84*, 88–96. [CrossRef]
6. Adams, K.; Heidarzadeh, M. A multi-hazard risk model with cascading failure pathways for the Dawlish (UK) railway using historical and contemporary data. *Int. J. Disaster Risk Reduct.* **2021**, *56*, 102082. [CrossRef]
7. Yadav, N.; Chatterjee, S.; Ganguly, A.R. Resilience of Urban Transport Network-of-Networks under Intense Flood Hazards Exacerbated by Targeted Attacks. *Sci. Rep.* **2020**, *10*, 10350. [CrossRef]
8. Baysari, M.T.; McIntosh, A.S.; Wilson, J.R. Understanding the human factors contribution to railway accidents and incidents in Australia. *Accid. Anal. Prev.* **2008**, *40*, 1750–1757. [CrossRef]
9. Wang, W.; Wang, Y.; Wang, G.; Li, M.; Jia, L. Identification of the critical accident causative factors in the urban rail transit system by complex network theory. *Phys. A Stat. Mech. Its Appl.* **2023**, *610*, 128404. [CrossRef]
10. Tang, H.; Zheng, J.; Li, M.; Shao, Z.; Li, L. Gauging the Evolution of Operational Risks for Urban Rail Transit Systems under Rainstorm Disasters. *Water* **2023**, *15*, 2811. [CrossRef]
11. Yu, K.; Wang, Y.; Zhao, W. Safety evaluation system of urban traffic network based on topological genetic algorithm. *J. Intell. Fuzzy Syst.* **2020**, *38*, 6825–6832. [CrossRef]
12. Lu, Q. Modeling network resilience of rail transit under operational incidents. *Transp. Res. Part A Policy Pract.* **2018**, *117*, 227–237. [CrossRef]
13. Liu, S.; You, Q.; Zhang, S. The study of subway operation safety management based on statistical analysis of accidents. *Saf. Secur.* **2019**, *40*, 62–66. [CrossRef]

14. Alliance, R. The Bad, Old Normal: Subway Signal Problems Still Delayed Trains During 80% of Morning Rush Hours in the First Half of 2021. Available online: https://static1.squarespace.com/static/61033b9bd377817f5bcc6db9/t/614e54a63c9104704b083efc/1635951289660/The+Bad,+Old+Normal:+Subway+Signal+Delays+-+Riders+Alliance (accessed on 25 July 2024).
15. The Paper. After Digging Through 5 Years of Data on the Beijing Subway, We Found the Most "Tragic" Commuter Route. Available online: https://www.thepaper.cn/newsDetail_forward_2676395 (accessed on 25 July 2024).
16. Esposito Amideo, A.; Starita, S.; Scaparra, M.P. Assessing Protection Strategies for Urban Rail Transit Systems: A Case-Study on the Central London Underground. *Sustainability* **2019**, *11*, 6322. [CrossRef]
17. Kiliç, O.; Çerçioğlu, H. Application of compromise multiple criteria decision making methods for evaluation of TCDD's railway lines projects. *J. Fac. Eng. Archit. Gazi Univ.* **2016**, *31*, 211–220.
18. Peng, P.; Liu, Z.; Guo, J.; Wang, C. Dynamic Metro Stations Importance Evaluation Based on Network Topology and Real-Time Passenger Flows. *KSCE J. Civ. Eng.* **2023**, *27*, 4459–4471. [CrossRef]
19. Xu, H.; Jiao, L.; Chen, S.; Deng, M.; Shen, N. An Innovative Approach to Determining High-Risk Nodes in a Complex Urban Rail Transit Station: A Perspective of Promoting Urban Sustainability. *Sustainability* **2018**, *10*, 2456. [CrossRef]
20. Xu, X.; Liu, J.; Li, H.; Hu, J. Analysis of subway station capacity with the use of queueing theory. *Transp. Res. Part C Emerg. Technol.* **2014**, *38*, 28–43. [CrossRef]
21. Zhang, X.; Zhang, W.; Lee, P.T.-W. Importance rankings of nodes in the China Railway Express network under the Belt and Road Initiative. *Transp. Res. Part A Policy Pract.* **2020**, *139*, 134–147. [CrossRef]
22. Meng, Y.; Tian, X.; Li, Z.; Zhou, W.; Zhou, Z.; Zhong, M. Exploring node importance evolution of weighted complex networks in urban rail transit. *Phys. A Stat. Mech. Its Appl.* **2020**, *558*, 124925. [CrossRef]
23. Li, X.; Zhang, P.; Zhu, G. Measuring Method of Node Importance of Urban Rail Network Based on H Index. *Appl. Sci.* **2019**, *9*, 5189. [CrossRef]
24. Meng, Y.; Qi, Q.; Liu, J.; Zhou, W. Dynamic Evolution Analysis of Complex Topology and Node Importance in Shenzhen Metro Network from 2004 to 2021. *Sustainability* **2022**, *14*, 7234. [CrossRef]
25. Starita, S.; Esposito Amideo, A.; Scaparra, M.P. Assessing Urban Rail Transit Systems Vulnerability: Metrics vs. Interdiction Models. In Proceedings of the Critical Information Infrastructures Security, Lucca, Italy, 8–13 October 2017; Springer: Cham, Switzerland, 2018; pp. 144–155.
26. Wollmer, R.D. Removing Arcs from a Network. *Oper. Res.* **1964**, *12*, 934–940. [CrossRef]
27. Fulkerson, D.R.; Harding, G.C. Maximizing the minimum source-sink path subject to a budget constraint. *Math. Program.* **1977**, *13*, 116–118. [CrossRef]
28. Li, J.; Lin, Y.; Su, Q. Identifying critical nodes of cyber–physical power systems based on improved adaptive differential evolution. *Electr. Power Syst. Res.* **2024**, *229*, 110112. [CrossRef]
29. Belkadi, K.; Lehsaini, M.; Tahraoui, M.A. Fault-tolerance based on augmenting approach in wireless sensor networks. *Concurr. Comput. Pract. Exp.* **2022**, *34*, e7359. [CrossRef]
30. Zhang, J.; Ren, G.; Song, J. Resilience-Based Restoration Sequence Optimization for Metro Networks: A Case Study in China. *J. Adv. Transp.* **2022**, *2022*, 22. [CrossRef]
31. Yu, W.; Wang, T.; Zheng, Y.; Chen, J. Parameter Selection and Evaluation of Robustness of Nanjing Metro Network Based on Supernetwork. *IEEE Access* **2019**, *7*, 70876–70890. [CrossRef]
32. Zhang, Y.; Ayyub, B.M.; Saadat, Y.; Zhang, D.; Huang, H. A double-weighted vulnerability assessment model for metrorail transit networks and its application in Shanghai metro. *Int. J. Crit. Infrastruct. Prot.* **2020**, *29*, 100358. [CrossRef]
33. Li, T.; Rong, L. A comprehensive method for the robustness assessment of high-speed rail network with operation data: A case in China. *Transp. Res. Part A Policy Pract.* **2020**, *132*, 666–681. [CrossRef]
34. Rodríguez-Núñez, E.; García-Palomares, J.C. Measuring the vulnerability of public transport networks. *J. Transp. Geogr.* **2014**, *35*, 50–63. [CrossRef]
35. Luo, C.; Xu, L. Railway disruption management: Designing bus bridging services under uncertainty. *Comput. Oper. Res.* **2021**, *131*, 105284. [CrossRef]
36. Yin, Y.; Huang, W.; Xie, A.; Li, H.; Gong, W.; Zhang, Y. Syncretic K-shell algorithm for node importance identification and invulnerability evaluation of urban rail transit network. *Appl. Math. Model.* **2023**, *120*, 400–419. [CrossRef]
37. Chen, T.; Ma, J.; Zhu, Z.; Guo, X. Evaluation Method for Node Importance of Urban Rail Network Considering Traffic Characteristics. *Sustainability* **2023**, *15*, 3582. [CrossRef]
38. Hu, J.; Zhang, W.; Tang, R.; Wen, C. A method of identifying critical stations for urban rail transit under bidirectional interruptions. *Ind. Eng. J.* **2023**, *26*, 168–174. [CrossRef]
39. Zhang, Y.; Zheng, S.; Chen, Y. Identification of Key Nodes in Comprehensive Transportation Network: A Case Study in Beijing-Tianjin-Hebei Urban Agglomeration, China. *Transp. Res. Rec.* **2023**, *2678*, 827–840. [CrossRef]
40. Zheng, Y.; Gao, L.; Chen, X.; Song, B.; Ding, L. Critical stations identification and robustness analysis of weighted metro-bus composite network. *J. Transp. Syst. Eng. Inf. Technol.* **2023**, *23*, 120–129. [CrossRef]
41. Zhou, W.; Li, T.; Ding, R.; Xiong, J.; Xu, Y.; Wang, F. Analysis of the Influence and Propagation Law of Urban Rail Transit Disruptions: A Case Study of Beijing Rail Transit. *Appl. Sci.* **2023**, *13*, 8040. [CrossRef]
42. Zhao, L.; Sun, P.; Zhang, J.; Peng, M.; Zhong, Y.; Liang, W. A Complex Network Important Node Identification Based on the KPDN Method. *Appl. Sci.* **2023**, *13*, 8303. [CrossRef]

43. Hu, J.; Yang, M.; Zhen, Y. A Review of Resilience Assessment and Recovery Strategies of Urban Rail Transit Networks. *Sustainability* **2024**, *16*, 6390. [CrossRef]
44. Zhang, W.; Hu, J.; Wen, C.; Tang, R. Research on urban rail failure recovery considering network resilience. *China Saf. Sci. J.* **2023**, *33*, 179–186. [CrossRef]
45. The Central People's Goverment of the People's Republic of China. Shanghai Subway"9.27" Accident Investigation Results Released 12 People Were Dealt with. Available online: https://www.gov.cn/gzdt/2011-10/07/content_1963474.htm (accessed on 25 July 2024).
46. Hernando, A.; Aguilera-Venegas, G.; Galán-García, J.L.; Nazary, S. An interlocking system determining the configuration of rail traffic control elements to ensure safety. *AIMS Math.* **2024**, *9*, 21471–21495. [CrossRef]
47. Chen, L.; Luo, G.; Ling, H. Positive significance for adopting ECC——Discussion on signal system extension for Shenzhen metro. *Mod. Urban Transit* **2010**, *2010*, 19–23. [CrossRef]
48. Taherdoost, H.; Madanchian, M. Multi-Criteria Decision Making (MCDM) Methods and Concepts. *Encyclopedia* **2023**, *3*, 77–87. [CrossRef]
49. Hajiagha, S.H.R.; Hashemi, S.S.; Mohammadi, Y.; Zavadskas, E.K.J.T. Fuzzy belief structure based VIKOR method: An application for ranking delay causes of Tehran metro system by FMEA criteria. *Transport* **2016**, *31*, 108–118. [CrossRef]
50. Jimu News. Nanjing Subway Morning Rush Hour Three Consecutive Failures, Passengers Received an Apology Letter, Official: Will Strengthen the Equipment Maintenance. Available online: https://new.qq.com/rain/a/20231106A07Y6N00 (accessed on 16 October 2024).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

sensors

MDPI

Article

A Study of Mixed Non-Motorized Traffic Flow Characteristics and Capacity Based on Multi-Source Video Data

Guobin Gu [1], Xin Sun [2], Benxiao Lou [1,3], Xiang Wang [4], Bingheng Yang [4], Jianqiu Chen [1,*], Dan Zhou [4], Shiqian Huang [4], Qingwei Hu [4] and Chun Bao [1]

[1] Guangxi Key Laboratory of International Join for China-ASEAN Comprehensive Transportation, Nanning University, Nanning 530200, China; guguobin@unn.edu.cn (G.G.); benxiaolou@csu.edu.cn (B.L.); 15007713136@163.com (C.B.)
[2] China Nuclear Engineering Consulting Co., Ltd., Beijing 102200, China; lynn@163.com
[3] School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China
[4] Guangxi Key Laboratory of Intelligent Transportation System (ITS), Guilin University of Electronic Technology, Guilin 541004, China; 17339200188@163.com (X.W.); 15726081096@163.com (B.Y.); 15251707205@163.com (D.Z.); hsq1216640923@outlook.com (S.H.); 199711@163.com (Q.H.)
* Correspondence: chenjianqiu@unn.edu.cn

Abstract: Mixed non-motorized traffic is largely unaffected by motor vehicle congestion, offering high accessibility and convenience, and thus serving as a primary mode of "last-mile" transportation in urban areas. To advance stochastic capacity estimation methods and provide reliable assessments of non-motorized roadway capacity, this study proposes a stochastic capacity estimation model based on power spectral analysis. The model treats discrete traffic flow data as a time-series signal and employs a stochastic signal parameter model to fit stochastic traffic flow patterns. Initially, UAVs and video cameras are used to capture videos of mixed non-motorized traffic flow. The video data were processed with an image detection algorithm based on the YOLO convolutional neural network and a video tracking algorithm using the DeepSORT multi-target tracking model, extracting data on traffic flow, density, speed, and rider characteristics. Then, the autocorrelation and partial autocorrelation functions of the signal are employed to distinguish among four classical stochastic signal parameter models. The model parameters are optimized by minimizing the AIC information criterion to identify the model with optimal fit. The fitted parametric models are analyzed by transforming them from the time domain to the frequency domain, and the power spectrum estimation model is then calculated. The experimental results show that the stochastic capacity model yields a pure EV capacity of 2060–3297 bikes/(h·m) and a pure bicycle capacity of 1538–2460 bikes/(h·m). The density–flow model calculates a pure EV capacity of 2349–2897 bikes/(h·m) and a pure bicycle capacity of 1753–2173 bikes/(h·m). The minimal difference between these estimates validates the effectiveness of the proposed model. These findings hold practical significance in addressing urban road congestion.

Keywords: traffic flow theory; mixed traffic flow; capacity; stochastic signals; power spectrum estimation

1. Introduction

Rapid urbanization and economic growth have led to a surge in urban populations and private vehicle numbers, intensifying urban traffic congestion and causing significant inconvenience for urban travel [1,2]. In response, governments worldwide have advocated for low-carbon, environmentally friendly solutions by promoting mixed non-motorized transportation modes. This shift encourages a transition from vehicle-centric travel to more efficient and eco-friendly modes, including public transit, rail systems, and mixed non-motorized vehicles [3–5]. Among these modes, mixed non-motorized transportation stands out as environmentally friendly, health-promoting, and convenient, helping to alleviate

urban congestion. They have already emerged as a primary a "last-mile" transportation solution in urban settings [6,7]. Consequently, mixed non-motorized vehicles have become integral to urban transportation systems, enhancing urban traffic conditions and serving as a strategic focus in modern urban transportation development.

Currently, severe urban traffic congestion in China has led to increased focus on the role of mixed non-motorized vehicles and pedestrians within integrated urban transportation systems. A comprehensive understanding of mixed non-motorized traffic flow characteristics and capacity within the urban system is needed. China's mixed non-motorized traffic primarily comprises human-powered and electric bicycles, with significant differences in their traffic flow characteristics. Their combined traffic flow characteristics are more complex than those observed in single-mode traffic [8–10]. In this study, mixed non-motorized traffic flow refers specifically to the traffic flow of human-powered and electric bicycles on bicycle lanes. Examining bicycle lane capacity holds both theoretical and practical significance for optimizing urban transportation, improving traffic efficiency, enhancing spatial planning, and bettering the travel experience of urban residents.

This article's main contribution is the proposal of a stochastic capacity estimation model based on power spectrum estimation, which simulates traffic flow characteristic parameters as stochastic signals and fits a stochastic signal parameter model, providing an improved representation of mixed non-motorized traffic flow dynamics. Additionally, changes in mixed non-motorized traffic flow parameters are analyzed in the frequency domain. The power spectrum estimation model is incorporated to estimate the non-motorized roadway capacity, revealing intrinsic relationships between spectral characteristics and traffic flow parameters. Finally, the collected data on stochastic passing capacity are used to verify the proposed model.

2. Literature Review

2.1. Study on Mixed Non-Motorized Traffic Flow Characteristics

This study investigates non-motorized traffic flow characteristics and mixed non-motorized lane capacity estimation, selecting three non-motorized lanes on main roads in Nanning and Hangzhou for data collection. A stochastic capacity estimation model based on power spectrum analysis is proposed. Therefore, exploring mixed non-motorized traffic flow characteristics, lane capacity, and passenger car equivalent metrics is necessary. Qin et al. [11] evaluated traffic efficiency concerning the speed characteristics of mixed non-motorized traffic flow. Their simulation of speed distribution and behavioral dynamics within mixed non-motorized traffic flow offers insights for urban management design. Hu et al. [12] examined speed characteristics of non-motorized traffic flow on a G310 national highway section in Henan Province, China, finding a close match between free-flow conditions and ideal speed distributions. Additionally, they identified an exponential relationship between flow rate and density and a cubic relationship between speed and distance. Chen et al. [13] employed mixed-model simulation to analyze traffic performance across different types of non-motorized vehicles under varying density conditions, aiming to optimize traffic efficiency in mixed-road settings. Alexander Bigazzi et al. [14] investigated the impact of electric bicycles on the capacity within bike lanes and mixed traffic settings, introducing the "Bicycle Equivalent Unit" (BEU) concept to measure space occupancy of electric bicycles relative to conventional bicycles, revealing that non-motorized traffic flow exhibits compressibility. Kieran Kalair et al. [15] analyzed high-resolution traffic data fluctuations on a bivariate probability density–flow plane and proposed an innovative real-time anomaly detection method. They found that electric bicycle speeds and flow rates are lower in non-congested conditions compared to congested ones.

In studying mixed non-motorized vehicle traffic flow characteristics, Jin et al. [16] proposed an improved Extended Burgers Cellular Automaton (EBCA) model for heterogeneous bicycle traffic flow, incorporating lane width-related compression probability. This model addresses limitations found in traditional models under various lane widths. Georgios Grigoropoulos [17] quantified traffic flow characteristics such as speed, acceleration,

and queue density in mixed non-motorized traffic at major signalized intersections, using the overtaking frequency as an indicator of vehicle interaction, which he found impacts traffic flow efficiency. Maria J et al. [18] introduced a distance-based speed function in a multi-class model to evaluate mixed non-motorized traffic flow, enabling travel time estimation and an overall road performance assessment. Feng et al. [19] examined the extended characteristics of left-turning non-motorized vehicles at intersections, establishing a multiple regression model to explore the relationship between extension degree and related factors, providing theoretical support for optimizing intersection traffic organization. Liu et al. [20] proposed a multi-agent model integrated with deep learning algorithms to simulate traffic conflicts between motor and non-motorized vehicles, accurately predicting evasive maneuvers and collision mechanisms with a strong correlation to actual trajectories.

Further advancements include Gu et al. [21], who developed a Dynamic Correlation Graph Convolutional Network (DCGCN) to construct adjacency matrices from input data using correlation coefficients. Li et al. [22] addressed parameter selection issues for bidirectional long short-term memory (BiLSTM) and kernel extreme learning machine (KELM) models, proposing SHO-BiLSTM and NOA-KELM, respectively, to enhance performance. Huan et al. [23] presented a genetic timing scheduling model (GTSM) for urban traffic signal control, using cellular automata to update timing cycles at multiple intersections, while Wen et al. [24] proposed an inner–outer cellular automaton combined with particle swarm optimization (IOCA-PSO) for real-time urban traffic light optimization.

2.2. Study on the Capacity of Mixed Non-Motorized Traffic

In studying bicycle capacity, Dilay Çelebi [25] proposed an integrated approach to optimize bicycle-sharing systems, combining location decision-making with capacity allocation to address unmet demand and balance system use. Navin [26] investigated the relationship between bicycle lane capacity and lane width in Canada, noting that a 2.5 m wide bicycle lane can accommodate up to 10,000 bicycles per hour. Converted to a 1 m wide lane, the capacity is approximately 4000 bicycles per hour. Botma [27] determined that a 1 m wide bicycle lane in the Netherlands has a capacity of 3200 bicycles per hour. Liu et al. [28] conducted a survey and found that Beijing's average bicycle crossing capacity is approximately 1600 bicycles per hour. Guo et al. [29] conducted field experiments to analyze bicycle traffic dynamics at bottlenecks, finding that capacity remained relatively stable after bottleneck activation. Matthew Brown et al. [30] analyzed GPS data and found that enhanced connectivity and physically separated infrastructure significantly increased bicycle capacity.

2.3. Passenger Car Equivalents (PCEs)

When analyzing traffic flows with varied vehicle types, it is essential to compare them on a standardized scale. In this context, traffic volumes of standard cars and various vehicle types are converted into passenger car equivalents (PCEs), an essential method in traffic analysis [31,32]. PCE denotes the number of standard cars equivalent to one vehicle of a specific type in terms of the traffic capacity.

Significant progress has been made in PCE study, resulting in various calculation methods. Classic PCE calculation methods include direct methods (e.g., mathematical models, capacity calculations, speed flow calculations), indirect methods (e.g., overtaking rate, delay calculation, headway calculation, leader vehicle method), and simulation methods [33,34].

Extensive studies have been conducted on conversion factors for electric bicycles relative to conventional bicycles. Given the limitations of existing PCE models in capturing the density characteristics of mixed non-motorized traffic, Zhang et al. [35] proposed a method to evaluate electric bicycle drive system efficiency, establishing a basis for conversion factors between electric and conventional bicycles. Wang et al. [36] developed a model to calculate the conversion factor using video analysis, yielding a factor of 1.3. Jaclyn S [37] analyzed the impact of electric bicycles on traffic flow and speed, finding a PCE

of 1.5 for electric bicycles in multimodal traffic, thus highlighting their greater efficiency in high-traffic settings compared to conventional bicycles. These studies indicate that conversion factors for electric bicycles relative to conventional bicycles converge around similar values.

2.4. Summary of Study Status

In summary, research on mixed non-motorized traffic flow characteristics, lane capacity, and passenger car equivalents (PCEs) offers theoretical support for the management, design, and evaluation of urban mixed non-motorized lanes. The current research can be summarized as follows:

(1) Current literature on mixed non-motorized traffic flow characteristics primarily describes parameters such as flow, density, and speed on urban roads, using mathematical methods like probability statistics for simulation and analysis.
(2) Studies on mixed non-motorized traffic capacity indicate that foreign researchers primarily use survey data and questionnaires to analyze lane capacity and service level, considering factors like lane width and capacity influences at intersections.

Domestic researchers focus on defining non-motorized lane capacity, mixed traffic flow capacity, lane geometry, and the impact of roadside facilities on capacity. Study methods include metacellular automata simulations and traffic conflict analysis, as well as studies examining conflicts between mixed non-motorized vehicles and passengers near bus stops.

Different road types (e.g., urban roads vs. highways), traffic density, and traffic composition influence PCE calculations. Vehicle type, weight, speed, traffic mobility, and environmental factors should also be considered in PCE calculations.

Existing studies also have some shortcomings:

(1) The stochastic nature of traffic flow changes limits the accuracy of methods like probability statistics, necessitating more complex models for precise description. Furthermore, modeling mixed non-motorized traffic flow often requires idealizing the operational environment, which fails to fully capture traffic flow dynamics.
(2) The current study primarily examines intersections or road sections, often focusing on external interference impacts on mixed non-motorized vehicle flow while neglecting stochastically triggered traffic conflicts during travel.

3. Modeling and Methodology

3.1. Data Acquisition Methods

3.1.1. YOLO-Based Image Detection

Detection area setup in aerial video: After aerial video collection, the detection area is defined in the video frame, with specific entrance and exit zones designated. When a vehicle enters the detection entrance, it is identified and assigned a unique number; upon exiting, its identification is removed. The detection area setup is illustrated in Figure 1.

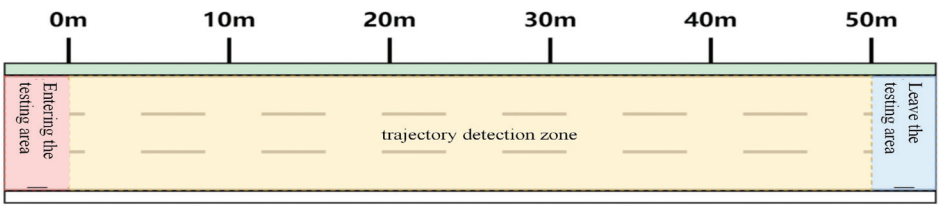


Figure 1. Schematic diagram of detection area.

The YOLO (You Only Look Once) algorithm [38] is employed for vehicle recognition in the captured video. YOLO's convolutional neural network (backbone network) divides the input image into equally sized cells without overlap. Each cell simultaneously participates in convolutional feature extraction, detecting the centroid of each feature layer within the cell. The model then outputs the probability of the target category's presence and the

target's predicted coordinates at that location [39]. In this study, YOLO is applied to the captured video to detect vehicles in each frame. The model identifies vehicles based on feature frame outputs, including centroid position, length, width, and other relevant data, enabling accurate identification of mixed non-motorized vehicles.

3.1.2. Mixed Non-Motorized Multi-Target Tracking

The DeepSORT multi-target tracking algorithm is utilized for tracking purposes. This tracking algorithm analyzes and extracts trajectory characteristics to address limitations of target detection, effectively removing false detections, enhancing accuracy, and supporting further behavioral analysis [40].

Target tracking involves locating specific targets across video frames, divided into two algorithmic stages. The first stage predicts the positions of targets in the Sth frame, projecting their appearance in the S + Nth frame (candidate region). The second stage verifies if the candidate region in the S + Nth frame corresponds to the initial targets in the Sth frame, initiating tracking if confirmed [41]. The process operates as follows: upon obtaining target detection results, the Kalman filter is initialized with the target frame. An eight-dimensional space represents the state of the track, depicting center position, aspect ratio, target frame height, and corresponding velocity information in image coordinates. This information is used to calculate positional and appearance feature relationships between the prediction frame and target detection frame, enabling association judgment between detection and tracking frames to complete multi-target tracking [42].

3.1.3. Virtual Coil Method

The virtual coil method is used to count traffic captured by the road test camera. At selected road survey points, two parallel lines are demarcated perpendicular to non-motorized road traffic, spaced 4 m apart, with the area between them designated as the virtual coil. This clamped region serves as the traffic video data detection area.

3.1.4. Artificial Assistance Method

The survey revealed that manual assistance is required to determine motor vehicle riders' gender and age. The questionnaire survey was conducted near the video data collection area, targeting riders passing through the study section. Video recording and the survey were conducted simultaneously.

3.2. Data Cleaning

3.2.1. Coordinate Conversion

UAV data collection at high altitudes is influenced by airflow, causing uniform rotational distortion in video frames. This study addresses the issue by repositioning the data origin to the center of the frame and applying coordinate rotation equations (Equations (1) and (2)) for data correction.

$$X_j = \sqrt{X_i^2 + Y_i^2} \cdot \cos(\arctan(\frac{Y_i}{X_i}) - \theta) \tag{1}$$

$$Y_j = \sqrt{X_i^2 + Y_i^2} \cdot \sin(\arctan(\frac{Y_i}{X_i}) - \theta) \tag{2}$$

where (X_j, Y_j) and (X_i, Y_i) denote the coordinate values before and after the center rotation process, respectively; θ denotes the angle of uniform rotation of this UAV per unit time.

3.2.2. Data Smoothing and Noise Reduction

During video recording, various factors can interfere with recognition and tracking, resulting in outliers in the extracted parameters of mixed non-motorized traffic flow. To address this, an exponential moving average is applied to perform three rounds of smoothing

and noise reduction on traffic flow speed and acceleration. This process is described by Equations (3)–(5) as follows:

$$S_1(\mathrm{t}) = \eta X(t) + (1-\eta) \cdot S_1(t-1) \tag{3}$$

$$S_2(\mathrm{t}) = \eta S_1(t) + (1-\eta) \cdot S_2(t \cdot 1) \tag{4}$$

$$S_3(t) = \eta S_2(t) + (1-\eta) \cdot S_3(t-1) \tag{5}$$

where $X_{(t)}$ denotes the actual position of the object at moment t; $S_i(t)$ denotes the ith exponential smoothing value at moment t; $S_i(t-1)$ denotes the ith exponential smoothing value at moment $t-1$; and η denotes the smoothing coefficient (generally taken as 0.6–0.9).

3.3. Traffic Flow Speed Distribution Model

3.3.1. Normal Distribution

The normal distribution model is commonly used to describe the operating speed of motorized vehicles and mixed non-motorized traffic flow. Its probability density function is represented by Equation (6):

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{6}$$

where μ represents the average speed of the mixed non-motorized traffic, and σ denotes the standard deviation of the traffic speed.

3.3.2. Log-Normal Distribution

A stochastic variable follows a log-normal distribution if its logarithm follows a normal distribution. The probability density function of the log-normal distribution is represented by Equation (7):

$$f(\mathrm{x}) = \frac{1}{\mathrm{x}\sqrt{2\pi}\sigma} e\left(-\frac{(\ln \mathrm{x} - \mu)^2}{2\sigma^2}\right) \tag{7}$$

where μ is the mean speed of the mixed non-motorized traffic, and σ is the standard deviation of the traffic speed.

3.3.3. Gamma Distribution

The gamma distribution is a continuous extension of the Poisson distribution defined over positive real numbers. Its probability density function is represented by Equation (8):

$$f(x, \beta, \alpha) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \ \mathrm{x} \geq 0 \tag{8}$$

where α is the shape parameter, β is the scale parameter, and $\Gamma(\alpha)$ is the gamma function, which is given by Equation (9):

$$\Gamma(a) = \int_0^{+\infty} t^{\alpha-1} e^{-t} dt \tag{9}$$

3.3.4. Weibull Distribution

The Weibull distribution is a continuous probability distribution, with its probability density function defined in Equation (10):

$$f(x) = \begin{cases} \frac{\eta}{\lambda}\left(\frac{x}{\lambda}\right) e^{-(x/2)^n}, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{10}$$

where $\lambda > 0$ is the scale parameter, and $\eta > 0$ is the shape parameter.

When describing speed, it is necessary to determine whether the speed distribution follows a specific distribution model. The Kolmogorov–Smirnov (K-S) test is commonly used for this purpose. The K-S test is a non-parametric method based on the cumulative distribution function. It compares the observed and expected frequencies to determine whether the sample probability distribution conforms to a specified theoretical distribution. To test whether data with n observations follow a known distribution $F_0(x)$, assuming the probability distribution of the observed data is $F(x)$, the formula is given by Equation (11).

$$F(x) = \frac{N}{n} \tag{11}$$

where N is the number of $x_i \leq x$.

According to the Glivenko–Cantelli theorem, when $n \to \infty$ occurs, the sample probability distribution $\hat{F}_n(x)$ converges uniformly to the population distribution $F(x)$ with probability 1. Therefore, the distance from $F(x)$ to $F_0(x)$ can be defined by Equation (12):

$$\mathrm{D} = \sup|\mathrm{F(x)} - \mathrm{F_0(x)}| \tag{12}$$

where $\sup(\cdot)$ is the maximum value function.

The magnitude of D is used to evaluate the goodness of fit to the population distribution $F_0(x)$. In the calculation process, if there are n observations, the statistic D_n can be used in place of D, and the expression for D_n is given by Equation (13):

$$\mathrm{D_n} = \max_{1 \leq \mathrm{i} \leq \mathrm{n}} \{\max(|\mathrm{F(x_i)} - \mathrm{F_0(x_i)}|, |\mathrm{F(x_{i-1})} - \mathrm{F_0(x_{i-1})}|)\} \tag{13}$$

According to the reference table, when the sample size exceeds 35, the critical value of D at a significance level of 0.05 is given by Equation (14):

$$D_{Critical} = \frac{1.36}{\sqrt{n}} \tag{14}$$

where n is the sample size.

If D_n exceeds the critical value, it indicates that the sample does not follow the specified known distribution $F_0(x)$; otherwise, it follows the specified distribution.

3.4. Stochastic Capacity Estimation Model

The fitting of the stochastic signal $x(n)$ can be represented using a system function model with a rational function. The expression is given by Equation (15):

$$x(n) = \sum_{i=0}^{q} b_i \omega(n-1) - \sum_{k=1}^{p} a_k x(n-k) \tag{15}$$

where $x(n)$ represents the studied stochastic signal, which, in this model, refers to an array of discrete traffic flow characteristic parameter observations arranged in chronological order; $\omega(n)$ represents a white noise sequence with a mean of 0 and a variance denoted as σ_ω^2; $\omega(n)$ needs to be solved during the stochastic signal solution process and can also be understood as the residual of the model during the corresponding order fitting process; and σ_ω^2 serves as the criterion for model fitting in this study and must be determined through actual computation. If it is constant in the final model calculation, it can be proven that the input signal can be described by the fitted model; b_l and a_k are real constants, while $a_0 = 1$ b_l, and a_k are parameters of multiple regular signals decomposed during stochastic signal fitting. k and l are the orders selected in the stochastic signal simulation model solution process, which can also be understood as the superposition of k or l regular signals.

In practice, appropriate parameter models are selected based on the numerical characteristics of the collected data. Stochastic signals can be approximately represented by an

autoregressive model, moving average model, autoregressive moving average model, or differencing integrated moving average autoregressive model.

3.4.1. Autoregressive Model

If a stochastic signal $\{x_n\}$ can be fitted into a model with the structure shown in Equation (16), it is called a bth-order autoregressive model (AR model, autoregressive model), abbreviated as $AR(p)$:

$$\begin{cases} x(n) = a_0 + a_1 x(n-1) + a_2 x(n-2) + \cdots + a_p x(n-p) + \omega(n) \\ a_p \neq 0 \\ E[\omega(n)] = 0,\ Var[\omega(n)] = \sigma_p^2,\ E[\omega(n)\omega(s)] = 0,\ s \neq n \\ E[x(n)x(s)] = 0,\ \forall s < n \end{cases} \tag{16}$$

where $x(n)$ is the stochastic signal to be analyzed, which represents the collected discrete traffic flow data arranged in chronological order; $a_1, a_2, \cdots, a_p$ is a real constant; $a_p \neq 0$; $\{\omega_n\}$ is a stationary white noise sequence with zero mean and variance σ_ω^2, which can also be interpreted as the residual between the fitted model and the actual observed values; and $a_1, a_2, \cdots, a_p$ and σ_ω^2 are the parameters to be estimated by the model.

3.4.2. Moving Average Model

If the stochastic signal $\{x_n\}$ is fitted into a model with the structure shown in Equation (17), it is called a qth-order moving average model (MA model, moving average model), abbreviated as $MA(q)$:

$$\begin{cases} x(n) = \mu + \omega(n) - b_1\omega(n-1) - b_2\omega(n-2) - \cdots - b_q\omega(n-q) \\ b_q \neq 0 \\ E[\omega(n)] = 0, Var[\omega(n)] = \sigma_{aa}^2, E[\omega(n)\omega(s)] = 0, s \neq n \end{cases} \tag{17}$$

where $x(n)$ is the stochastic signal to be analyzed, $b_1, b_2, \cdots, b_q$ is a real constant, and $b_q \neq 0$; $\{\omega_n\}$ is a stationary white noise sequence with zero mean and variance σ_ω^2, representing the residual between the fitted model and the actual observed values. $\mu = \frac{\alpha_0}{1-\alpha_1-\cdots-\alpha_p}$ is a constant; $b_1, b_2, \cdots, b_q$ and σ_ω^2 are the parameters to be estimated by the model.

3.4.3. Autoregressive Moving Average Model

If the fitting form of the stochastic signal $\{x_n\}$ follows the structure shown in Equation (18), it is called an Autoregressive Moving Average model (ARMA model), abbreviated as $ARMA(p, q)$.

$$\begin{cases} x(n) = a_0 + a_1 x(n-1) + \cdots + a_p x(n-p) + \\ \omega(n) - b_1\omega(n-1) - \cdots - b_q\omega(n-q) \\ a_p \neq 0, b_q \neq 0 \\ E[\omega(n)] = 0, Var[\omega(n)] = \sigma_\omega^2, E[\omega(n)\omega(s)] = 0, s \neq n \\ E[x(s)\omega(n)] = 0, \forall s < n \end{cases} \tag{18}$$

where $x(n)$ is the stochastic signal to be analyzed, $a_1, a_2, \cdots, a_p$ and $b_1, b_2, \cdots, b_q$ are real constants, $a_p \neq 0$, $b_q \neq 0$; $\{\omega_n\}$ is a stationary white noise sequence with zero mean and variance σ_ω^2; and $a_1, a_2, \cdots, a_p$, $b_1, b_2, \cdots, b_q$, and σ_ω^2 are the parameters to be estimated by the model.

The ARMA model combines both the AR model and the MA model, and in parameter estimation, it is necessary to estimate both $a_1, a_2, \cdots, a_p$ and $b_1, b_2, \cdots, b_q$ simultaneously.

3.4.4. Differencing Integrated Moving Average Autoregressive Model

The ARMA model is one of the key methods in parameter estimation, offering higher accuracy in signal fitting than the AR and MA models. However, AR, MA, and ARMA models are limited to fitting stationary signals. In practice, most collected signals are

non-stationary, necessitating an extension of the ARMA model to represent non-stationary stochastic signals. This extension is known as the Autoregressive Integrated Moving Average (ARIMA) model. The ARIMA model incorporates differencing of any order with the ARMA model to describe non-stationary stochastic signals.

3.5. Power Spectrum Estimation Model

After applying the Z-transform to $x(n) = \sum_{i=0}^{p} b_i \omega(n-1) - \sum_{k=1}^{p} a_k x(n-k)$, the expression becomes Equation (19).

$$\sum_{k=1}^{p} a_k x(z) z^{-k} = \sum_{i=0}^{p} b_i \omega(z) z^{-1} \tag{19}$$

Let $A(z) = \sum_{k=1}^{p} a_k z^{-k}$ and $B(z) = \sum_{l=0}^{q} b_l z^{-1}$.

Then, the transfer function of the stochastic signal $x(n)$ after the Z-transform is given by Equation (20):

$$H(z) = \frac{X(z)}{W(z)} = \frac{\sum_{l=0}^{q} b_l z^{-1}}{\sum_{k=1}^{p} a_k z^{-k}} \triangleq \frac{B(z)}{A(z)} \tag{20}$$

The power spectral density $P_{xx}(z)$ of the parametric model is given by Equation (21):

$$p_\omega(z) = \sigma_\omega^2 H(z) H(z^{-1}) = \sigma_\omega^2 \frac{B(z) \cdot B(z^{-1})}{A(z) \cdot A(z^{-1})} \tag{21}$$

4. Experiments

4.1. Data Collection

Traffic flow characterization relies on accurate data; this study focuses on mixed non-motorized traffic flows, which have more complex operational characteristics than motorized traffic flows. The variety of vehicle types, sizes, and speeds on urban non-motorized lanes, along with the need for clear delineation from motorized lanes, introduces more stochasticness into mixed non-motorized traffic flows, complicating data collection. Thus, designing a scientifically sound survey program is essential, including selecting appropriate survey locations, acquiring relevant data types, and pre-processing the data to analyze mixed non-motorized traffic flows.

In this study, only the road width differs between Nanning and Hangzhou, while other road conditions are consistent; three types of non-motorized roads are selected for investigation, with UAVs primarily used to collect traffic flow video data. All survey locations are over 50 m from intersections, bus stops, pedestrian crossings, entrances, and other disturbances, with good road alignment and no slopes, curves, or unusual road forms. Additionally, non-motorized lanes are physically separated from motorized lanes by guardrails, ensuring that mixed non-motorized traffic is unaffected by other factors [43].

Due to varying road conditions across cities, this study adopts two data collection methods. In Nanning, UAVs capture footage, with the YOLO image detection algorithm identifying mixed non-motorized vehicles in the video, and the DeepSORT multi-target tracking algorithm tracking targets and calculates traffic flow data. In Hangzhou, roadside cameras were used for data collection, and the virtual coil method was applied for traffic counting.

In this study, the extracted mixed non-motorized vehicle data fields from the multi-source video include the number, entry and exit times in the detection zone, time difference, vehicle speed (km/h), type, gender, age, and remarks. Type, gender, and age were coded as follows: type (1 for human-powered bicycles, 2 for electric bicycles), gender (1 for males,

0 for females), and age (<29 for young people, 30–49 for middle-aged, and ≥50 for older adults). The basic information for each surveyed road is provided in Table 1.

Table 1. Basic information of surveyed road sections.

Cities	Sites	Breadth	Abbreviations
Nanning	Minzu Avenue (next to Shimen Forest Park)	3.55 m	Section A
	Minzu Avenue and Qingxiu Road East Section	3 m	Section B
	Yongjiang Bridge Road Section	2.2 m	Section C
Hangzhou	Dongxin-Shaoxing junction, Dongxin road section	3.65 m	Section D
	Hushu South-Chaowang Intersection	2.93 m	Section E
	Teachers-Wen Er Road junction	2.27 m	Section F

Figure 2 presents an actual map of each surveyed road section. The surveyed area includes flexible and non-rigid separation facilities, preventing motor vehicle intrusion and influence. The road surface is smooth, with clear markings, ensuring optimal visual conditions for drivers.

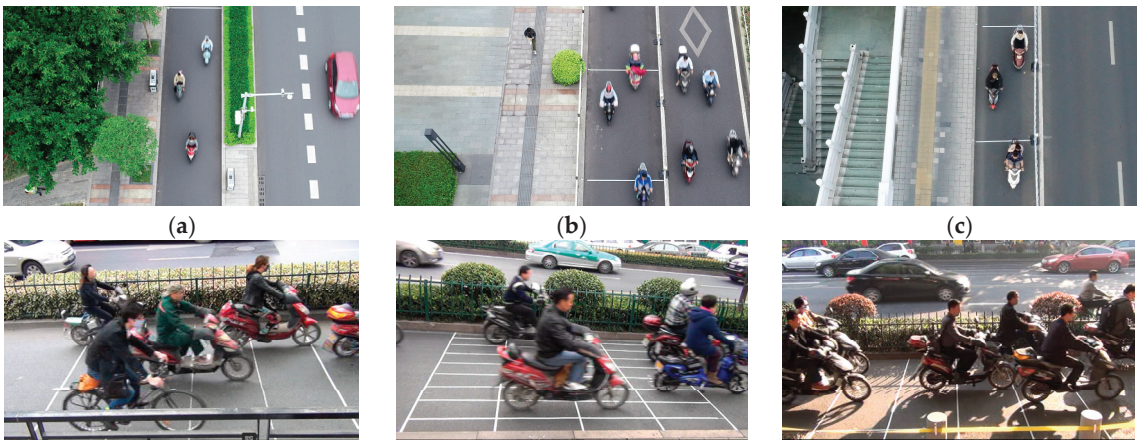


Figure 2. Live view of surveyed roadway. (**a**) Road section A. (**b**) Road section B. (**c**) Road section C. (**d**) Road section D. (**e**) Road section E. (**f**) Road section F.

Table 2 presents the age and gender statistics of cyclists collected through the questionnaire near each section, while Table 3 provides age and gender statistics obtained through video analysis.

Table 2. Statistical data table on age and gender for the questionnaire survey.

Cities	Sites	Percentage of Age			Percentage by Gender	
		0–29 Years	30–49 Years	50 Years and Over	Male	Female
Nanning	Minzu Avenue (next to Shimen Forest Park)	24.87%	71.48%	3.65%	72.68%	27.32%
	Minzu Avenue and Qingxiu Road East Section	23.37%	72.79%	3.84%	74.26%	25.74%
	Yongjiang Bridge Road Section	30.35%	66.57%	3.08%	76.41%	23.59%
Hangzhou	Dongxin-Shaoxing junction, Dongxin road section	52.10%	44.72%	3.17%	67.41%	32.59%
	Hushu South-Chaowang Intersection	45.59%	52.16%	2.25%	65.18%	34.82%
	Teachers-Wen Er Road junction	56.97%	40.69%	2.34%	68.17%	31.83%

Table 3. Statistical data table on age and gender for video discrimination.

Cities	Sites	Percentage of Age			Percentage by Gender	
		0–29 Years	30–49 Years	50 Years and Over	Male	Female
Nanning	Minzu Avenue (next to Shimen Forest Park)	21.63%	73.51%	4.86%	66.97%	33.03%
	Minzu Avenue and Qingxiu Road East Section	20.50%	74.51%	4.99%	68.31%	31.69%
	Yongjiang Bridge Road Section	25.94%	70.01%	4.05%	73.50%	26.50%
Hangzhou	Dongxin-Shaoxing junction, Dongxin road section	46.52%	48.88%	4.60%	58.52%	41.48%
	Hushu South-Chaowang Intersection	42.21%	55.14%	2.65%	59.57%	40.43%
	Teachers-Wen Er Road junction	53.25%	43.15%	3.60%	65.57%	34.43%

Calculating the error in age and gender ratios using the Euclidean distance shows that the difference between the two methods is relatively small, all within 0.1, indicating that age and gender data obtained via video analysis are reliable. The questionnaire and video analysis results overlap, making both methods valuable for cross-referencing. Part of the error may arise because the questionnaire survey is a sampling method, meaning respondents near the surveyed section may not fully represent all non-motorized cyclists on the road.

4.2. Traffic Flow Characteristics Analysis

4.2.1. Statistical Characteristics Analysis

A statistical analysis of traffic flow data uses observational data to examine the essential characteristics and underlying traffic flow patterns. This is a fundamental method for studying traffic flow issues. An analysis of mixed traffic flow characteristics includes factors such as types of mixed non-motorized vehicles, rider characteristics, traffic flow volume, and vehicle speed [44].

In this study, video image detection and object tracking algorithms are used to obtain video parameters of mixed non-motorized vehicles. The time when each non-motorized vehicle enters the detection area is recorded as t_1, and the time when it leaves the detection area is recorded as t_2. The length of the detection area is L = 15 m, and the speed of the non-motorized vehicle is denoted as v, which is calculated using Equation (22).

$$v = \frac{L}{t_2 - t_1} \tag{22}$$

In the equation, v represents the speed of the non-motorized vehicle; t_1 and t_2 are the times when the vehicle enters and leaves the detection area, respectively; and L is the length of the detection area.

Because non-motorized vehicle lanes lack strict lane divisions, area density is used to represent mixed non-motorized vehicle density. Defined as the number of mixed non-motorized vehicles per unit area at a given time, it is denoted by K, with the formula provided in Equation (23):

$$k = \frac{N}{L \times W} = \frac{N}{L \times (W - 0.5)} \tag{23}$$

In the equation, N is the number of mixed non-motorized vehicles within the observation range at a given time; L is the observation length; w is the effective width of the non-motorized vehicle lane; and W is the width of the non-motorized vehicle lane.

Study indicates that varying sampling time intervals can significantly impact the predictions of non-motorized vehicle lane capacity [45]. Given the short saturation time of mixed non-motorized traffic flow, a 30 s sampling interval was used. Based on the literature review in Section 2, which summarizes current research on passenger car equivalents, this study uses a conversion factor of 1.34 for electric bicycles relative to traditional bicycles, which was derived from previous research.

A statistical description of the survey data is provided in Table 4.

Table 4. Statistical description of road section survey data.

Cities	Investigated Road Sections	Lane Width	Sample Size	Percentage of Electric Vehicles	Percentage of Males	Proportion of Youth
Nanning	Minzu Avenue (next to Shimen Forest Park)	3.55 m	16,268	98.36%	66.97%	21.63%
	Minzu Avenue and Qingxiu Road East Section	3 m	10,076	97.33%	68.31%	20.50%
	Yongjiang Bridge Road Section	2.2 m	8979	96.74%	73.50%	25.94%
Hangzhou	Dongxin-Shaoxing junction, Dongxin road section	3.65 m	10,037	77.28%	58.52%	46.52%
	Hushu South-Chaowang Intersection	2.93 m	10,782	70.26%	59.57%	42.21%
	Teachers-Wen Er Road junction	2.27 m	12,515	66.44%	65.57%	53.25%
	average value	2.93 m	11,442	84.40%	65.41%	35.01%

Table 4 shows that traffic volume in the surveyed sections is relatively high, providing a sufficient sample size. A small number of drivers were observed traveling in the wrong direction. Due to the low incidence of wrong-way driving, its impact is considered minimal in this study.

Figure 3 presents the rider characteristics by gender, age, and vehicle type. In Figure 3a, the average proportion of male riders across the six surveyed sections is significantly higher than that of female riders, with males accounting for approximately 60%. This suggests male riders are more likely to use mixed non-motorized vehicles and may have a higher demand for speed, making them more inclined to choose electric bicycles over traditional bicycles. Figure 3b shows that most mixed non-motorized vehicle riders are middle-aged, though there are notable differences in age distribution between the two cities. In Nanning, riders aged 0–29 constitute 23.28%, lower than 42.85% in Hangzhou, suggesting regional differences in mixed non-motorized traffic flow. In Nanning's three surveyed sections, mixed non-motorized vehicle proportions are 98.36%, 97.33%, and 96.74%, while in Hangzhou, they are 77.28%, 70.26%, and 66.44%. Figure 3c also highlights significant differences in mixed non-motorized vehicle proportions between the two cities, suggesting that regional culture and economic development shape mixed non-motorized traffic flow characteristics. In summary, the analysis methods for mixed non-motorized traffic flow should be tailored to each city, underscoring the need to study the characteristics of emerging mixed non-motorized traffic flows as examined in this paper.

Figure 3 compares the average speeds of mixed non-motorized vehicles across six sections by vehicle type, rider gender, and rider age. The significant variation in average speeds across different types of mixed non-motorized vehicles is primarily due to performance differences between electric and traditional bicycles. Additionally, male and middle-aged riders exhibit higher average speeds across all sections, consistent with prior study findings [46]. Moreover, average speeds between the two cities vary significantly across age groups, influenced by regional culture, transportation infrastructure, and management policies.

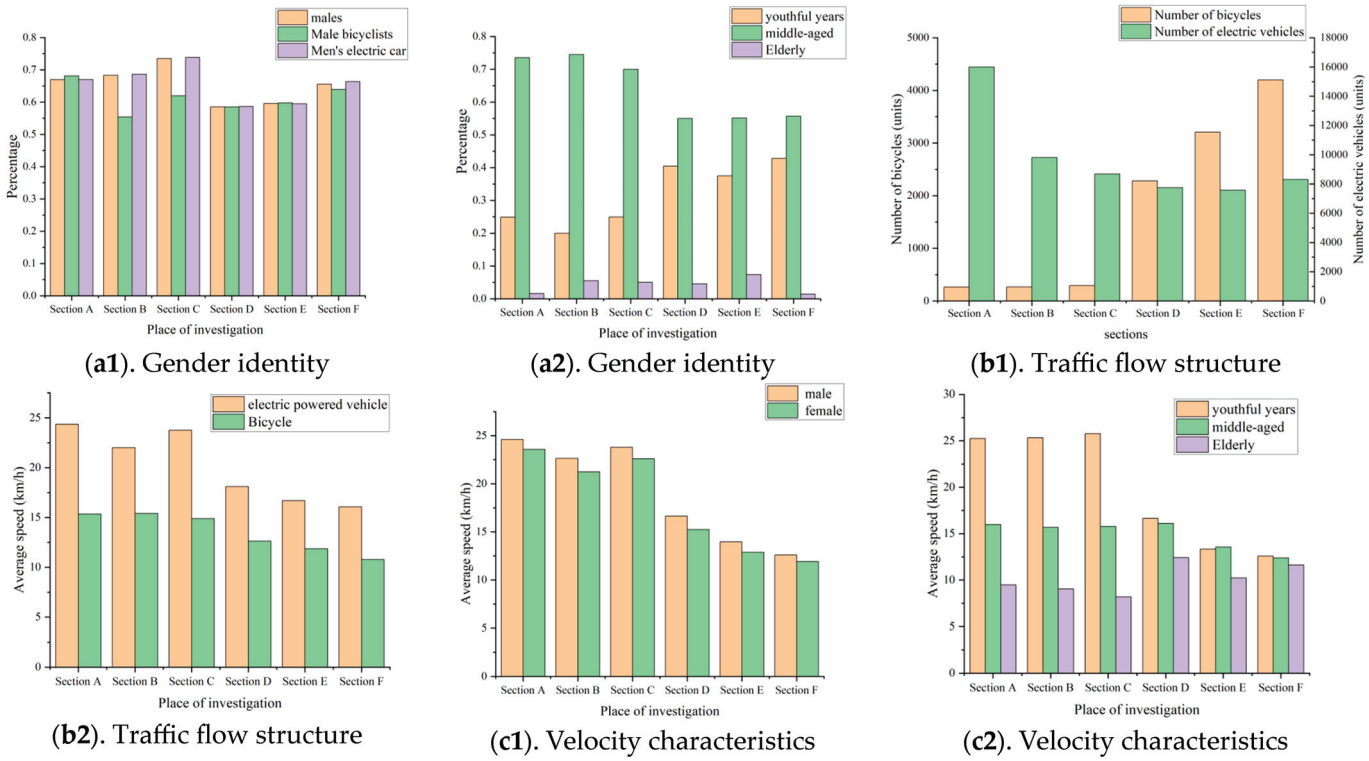


Figure 3. Characteristics of traffic flow statistics on the investigated roadway. (**a1**). Gender identity. (**a2**). Gender identity. (**b1**). Traffic flow structure. (**b2**). Traffic flow structure. (**c1**). Velocity characteristics. (**c2**). Velocity characteristics.

4.2.2. Traffic Flow Speed Distribution Analysis

Existing studies typically describe the probability distribution of mixed non-motorized vehicle operating speeds using a normal distribution. However, given the complexity of mixed non-motorized traffic flows and varying speed distributions across sections, researchers suggest using log-normal, gamma, and Weibull distributions to better represent these speeds [47]. Based on current findings, this study applies normal, log-normal, gamma, and Weibull models to describe the operating speeds of mixed non-motorized traffic across sections. These models fit and test the speed frequency distribution for each section, with results shown in Figure 4.

As shown in the speed distribution fitting diagram, the speed distribution of mixed non-motorized vehicles is dispersed due to significant speed differences among various vehicle types. This leads to interference and frequent conflicts between non-motorized vehicles, significantly impacting safety. For example, electric vehicles can reach speeds up to 56 km/h, exceeding the national limit of 25 km/h and posing a hidden risk to traffic safety.

The K-S test indicates a relatively poor fit for the distribution model when applied to mixed speeds of each road section. This suggests that mixed non-motorized vehicle flow speeds are influenced by multiple factors, adding stochasticness and complexity to their distribution characteristics. Traditional single-distribution models cannot adequately describe the speed distribution of mixed non-motorized vehicles.

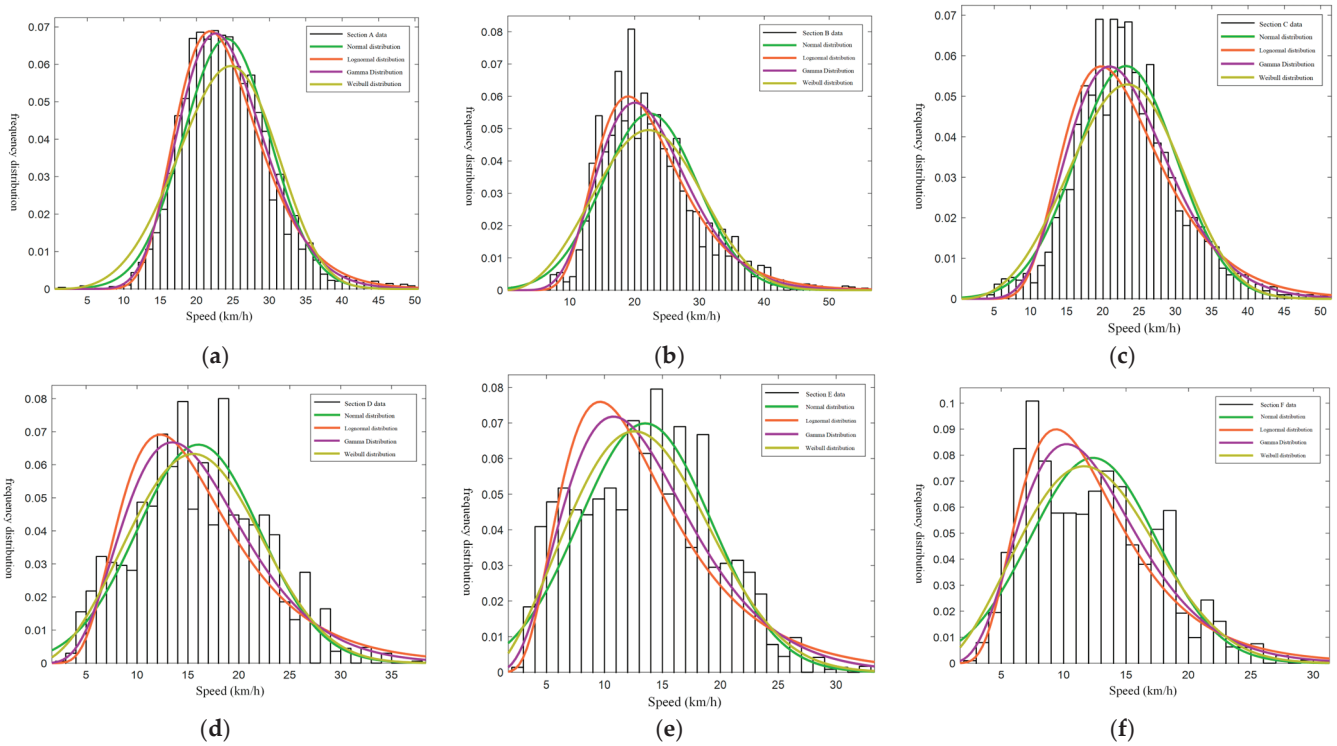


Figure 4. Speed distribution of surveyed road sections. (**a**) Segment A speed profiles. (**b**) Segment B speed profiles. (**c**) Segment C speed profiles. (**d**) Segment D speed profiles. (**e**) Segment E speed profiles. (**f**) Segment F speed profiles.

4.2.3. Stochastic Passability Estimation Based on Power Spectrum Estimation

The power spectrum estimation model reveals intrinsic patterns of stochastic signal variation in the frequency domain. The central petal in the power spectrum represents the highest peak, reflecting the intensity of the dominant frequency component, obtained after stochastic factors are eliminated. This peak, or central petal, indicates the passing capacity. The stochastic passing capacity model, based on power spectrum estimation, follows the steps shown in Figure 5:

Step 1: Smoothness analysis: Extract traffic flow data as a time-varying stochastic signal, and test the signal for smoothness to verify if the time series is stationary;

Step 2: Differencing: If the plotting and ADF tests indicate the time series is non-stationary, apply differencing to obtain a stationary sequence;

Step 3: Model identification: Determine model parameters based on autocorrelation and partial autocorrelation properties to select an appropriate range;

Step 4: Model optimization: Use the AIC and BIC criteria to find the optimal parameter combination and test for statistical significance;

Step 5: Model testing: Conduct hypothesis testing to determine if the sequence is white noise and evaluate model superiority;

Step 6: Calculate passing capacity: Extract the main peak intensity to determine the passing capacity of the road section.

This study selected the following road sections for experiments, with their primary conditions listed in Table 5. After data preprocessing, 2846 traffic flow observations were collected, using a 30 s statistical interval. The collected traffic flow video captures high-density traffic, including free-flowing conditions and severely congested periods. These three road sections were selected for stochastic signal analysis, allowing for the examination of each traffic state as it traverses. This analysis can reveal the underlying patterns of traffic flow characteristics within each road section.

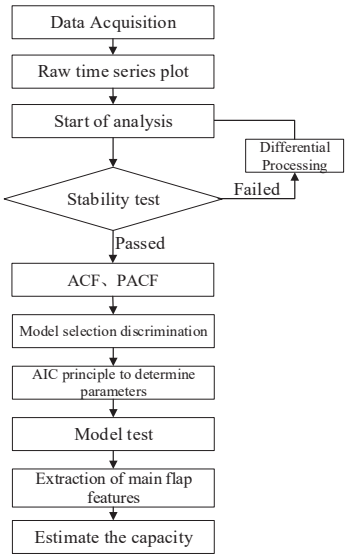


Figure 5. Flowchart for modeling the stochastic capacity model.

Table 5. Basic information of selected surveyed road sections.

Cities	Sites	Width	Abbreviations
Nanning	Minzu Avenue (next to Shimen Forest Park)	3.55 m	Section A
	Minzu Avenue and Qingxiu Road East Section	3 m	Section B
	Yongjiang Bridge Road Section	2.2 m	Section C
Hangzhou	Dongxin-Shaoxing junction, Dongxin road section	3.65 m	Section D
	Hushu South-Chaowang Intersection	2.93 m	Section E
	Teachers-Wen Er Road junction	2.27 m	Section F

After processing the traffic flow characteristics of the selected road sections, the traffic flow rate was plotted in Figure 6, using the sampling interval as a reference. Figure 6 illustrates the relationship between time and traffic flow rate across the surveyed sections. Each increment on the time axis represents a 30 s interval.

The stationarity of the signal can be preliminarily observed in Figure 6. The figure shows that the flow time series curve for section D fluctuates around a horizontal line, indicating that the mean of the flow time series signal is relatively stable, with minimal changes in the amplitude over time. This suggests that the variance is also stable, indicating that the flow stochastic signal for section D can be considered generally stationary. Similarly, the speed time series graphs for sections E and F also suggest stationary signals. Further testing is needed for signals without an observable trend in the time series graph. The primary method for stationarity testing is the unit root test, which examines whether a unit root exists in the sequence. If a unit root is present, the sequence is non-stationary; if it is absent, the sequence is stationary. Stationarity test results are presented in Table 6.

Table 6 presents the test statistic and corresponding p-value for each data collection, calculated through the stationarity test. The null hypothesis assumes the presence of a unit root, indicating a non-stationary time series. By testing signal stationarity and comparing probability values with significance levels of 1%, 5%, and 10%, it was found that the null hypothesis is not rejected for flow rates in sections A and B, indicating non-stationarity. The flow rates in sections C and F, with test statistics of -2.811 and -2.843 and p-values of 0.057 and 0.052, respectively, are below the 10% significance level, indicating stationary signals. The flow rate in section D, with a p-value of 0.035 at the 5% significance level, rejects the null hypothesis, indicating stationarity. The flow rate in section E, with a p-value of 0.001 at the 1% significance level, also rejects the null hypothesis, confirming stationarity. To meet stationarity requirements for power spectrum modeling, differencing must be applied to non-stationary sequences.

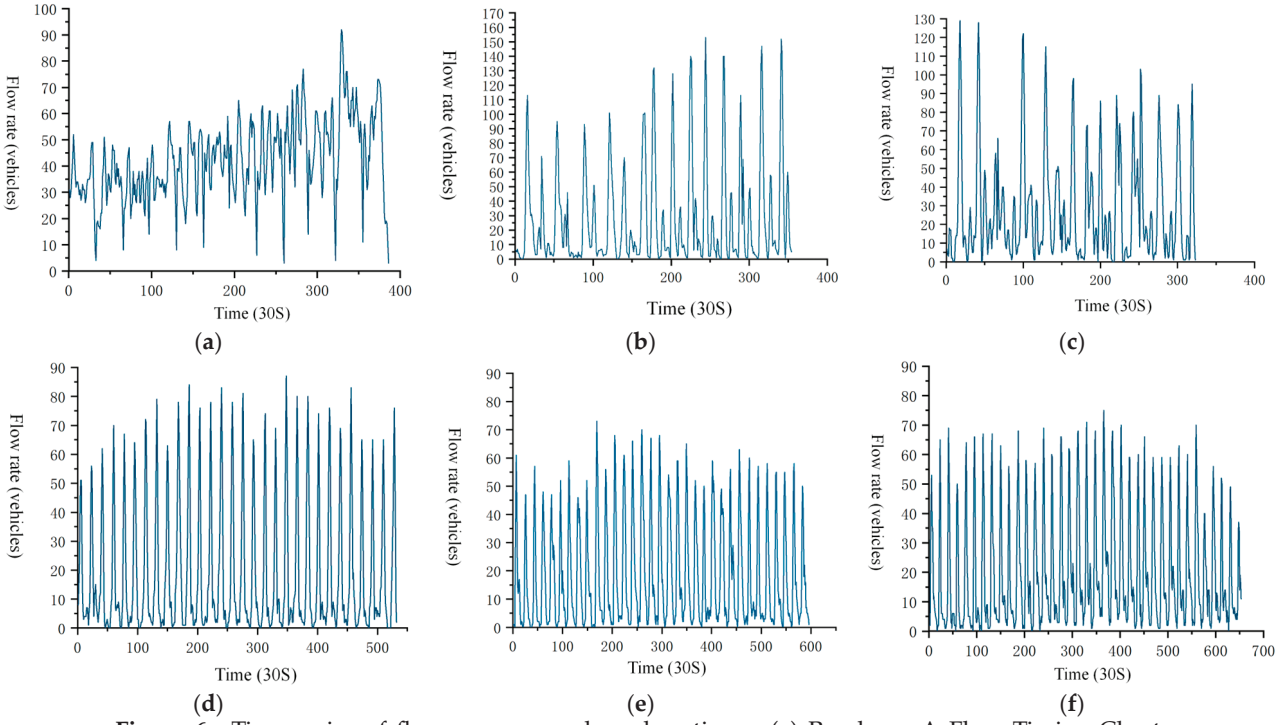


Figure 6. Time series of flow on surveyed road sections. (**a**) Roadway A Flow Timing Chart. (**b**) Roadway B Flow Timing Chart. (**c**) Roadway C Flow Timing Chart. (**d**) Roadway D Flow Timing Chart. (**e**) Roadway E Flow Timing Chart. (**f**) Roadway F Flow Timing Chart.

Table 6. Smoothness test table.

Data	Test Statistic	*p*-Value	1%	5%	10%	H
Section A Flow	−2.188	0.027	−3.466	−2.877	−2.575	0
section B Flow	−1.251	0.651	−3.471	2.88	−2.576	0
Section C Flow	−2.811	0.057	−3.47	−2.879	−2.576	1
Section D Flow	−3.002	0.035	−3.474	−2.879	−2.576	1
Section E Flow	−4.941	0.001	−3.465	−2.877	−2.575	1
Section F Flow	−2.843	0.052	−3.472	−2.88	−2.576	1

After first-order differencing on initially non-stationary signals, stationarity testing confirmed that the signals became stationary.

To reduce the computational load for the stochastic capacity model, this study performed autocorrelation and partial autocorrelation analyses to preliminarily determine the p and q orders. Figure 7 presents the autocorrelation and partial autocorrelation function graphs for traffic flow signals in sections A, B, and C.

In the power spectrum estimation model, the autocorrelation and partial autocorrelation function graphs exhibit two characteristics: tailing and cut-off. These characteristics can be used for preliminary testing to identify the presence of p-lags or q-lags. Figure 7 shows a clear geometric decay in the autocorrelation functions of the three sections, indicating p-lags in their flow signals. The partial autocorrelation function graphs for the three sections consistently show non-zero values, suggesting q-lags in all three signals. By analyzing the autocorrelation and partial autocorrelation functions of the selected section flow signals within a 95% confidence interval, the *p*- and q-values for the power spectrum estimation model were determined, as shown in Table 7. Here, d represents the model's differencing order.

The AIC information criterion was used to further refine the power spectrum model parameters and obtain optimal values. The residuals of the fitted model were tested using Q-statistic values, followed by a goodness-of-fit test. Using road section A as an example, Table 8 presents the flow model test results for section A.

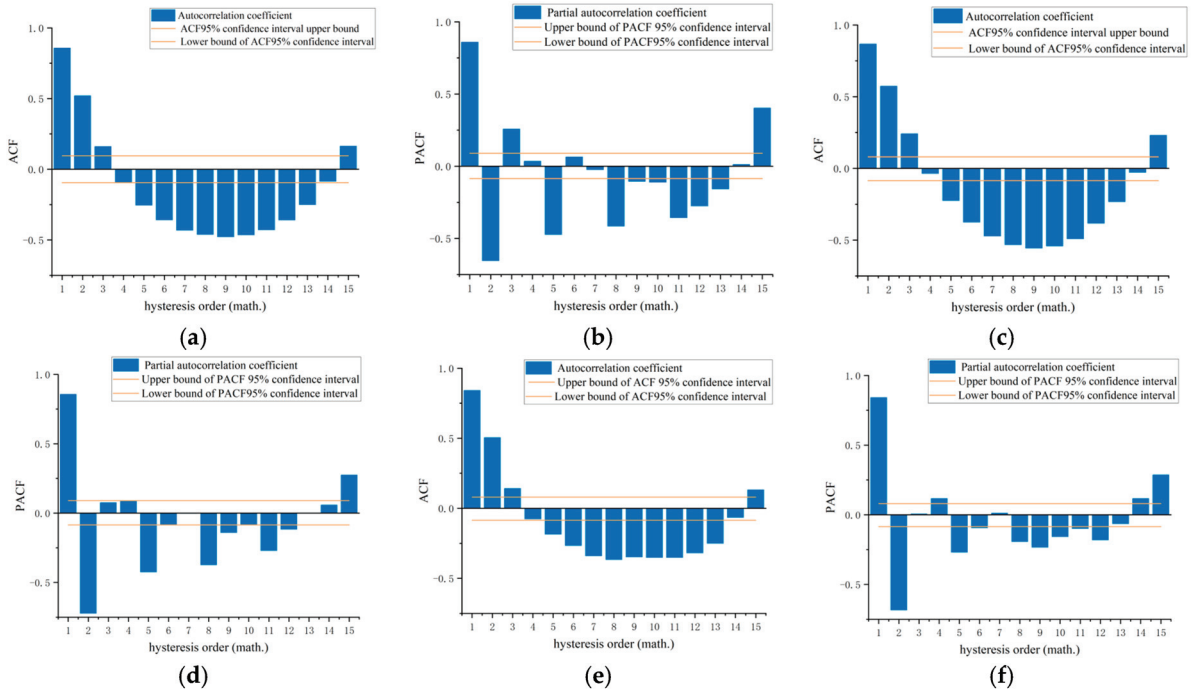


Figure 7. Images of flow signals on surveyed roadway segments. (**a**) Autocorrelation images of road segment C flows. (**b**) Segment A flow-biased autocorrelation image. (**c**) Autocorrelation images of road segment B flows. (**d**) Segment B flow biased autocorrelation image. (**e**) Autocorrelation images of road segment C flows. (**f**) Segment C flow biased autocorrelation image.

Table 7. Table of parameter ranges for power spectrum estimation.

Data	Model Parameter		
	p	d	q
Flow rate of Section A	4	1	4
Flow rate of Section B	4	1	5
Flow rate of section C	3	0	3
Flow rate of section D	4	0	4
Flow rate of section E	4	0	3
Flow rate of section F	4	0	3

Table 8. Flow modeling test table for section A.

ARIMA Model (3, 1, 2) Test Table			
Term	**Symbol**	**Value**	***p*-Value**
Q-Statistics	Q6 (*p*-value)	1.533	0.957
	Q12 (*p*-value)	12.635	0.068
	Q18 (*p*-value)	50.932	0.000
	Q24 (*p*-value)	55.897	0.000
	Q30 (*p*-value)	56.955	0.000
Information criterion	AIC	1927.95	
Goodness of fit	R^2	0.879	

Table 8 shows that the optimal parameters for the flow model of road section A, based on the AIC criterion, are ARIMA (3, 1, 2). Additionally, the optimal models for each section are as follows: road section B—ARIMA (2, 1, 3); road section C—ARIMA (2, 0, 3); road section D—ARIMA (2, 0, 4); road section E—ARIMA (2, 0, 2); and road section F—ARIMA (2, 0, 2).

Figure 8 presents the power and frequency spectra of traffic flow for the selected section. In the frequency spectrum, the component with the highest intensity is marked by

a red box, representing the signal with the greatest amplitude variation derived from the stochastic signal. The corresponding frequency in the power spectrum, also highlighted by a red box, represents the capacity sought by this model, indicated by the peak value of the most prominent component.

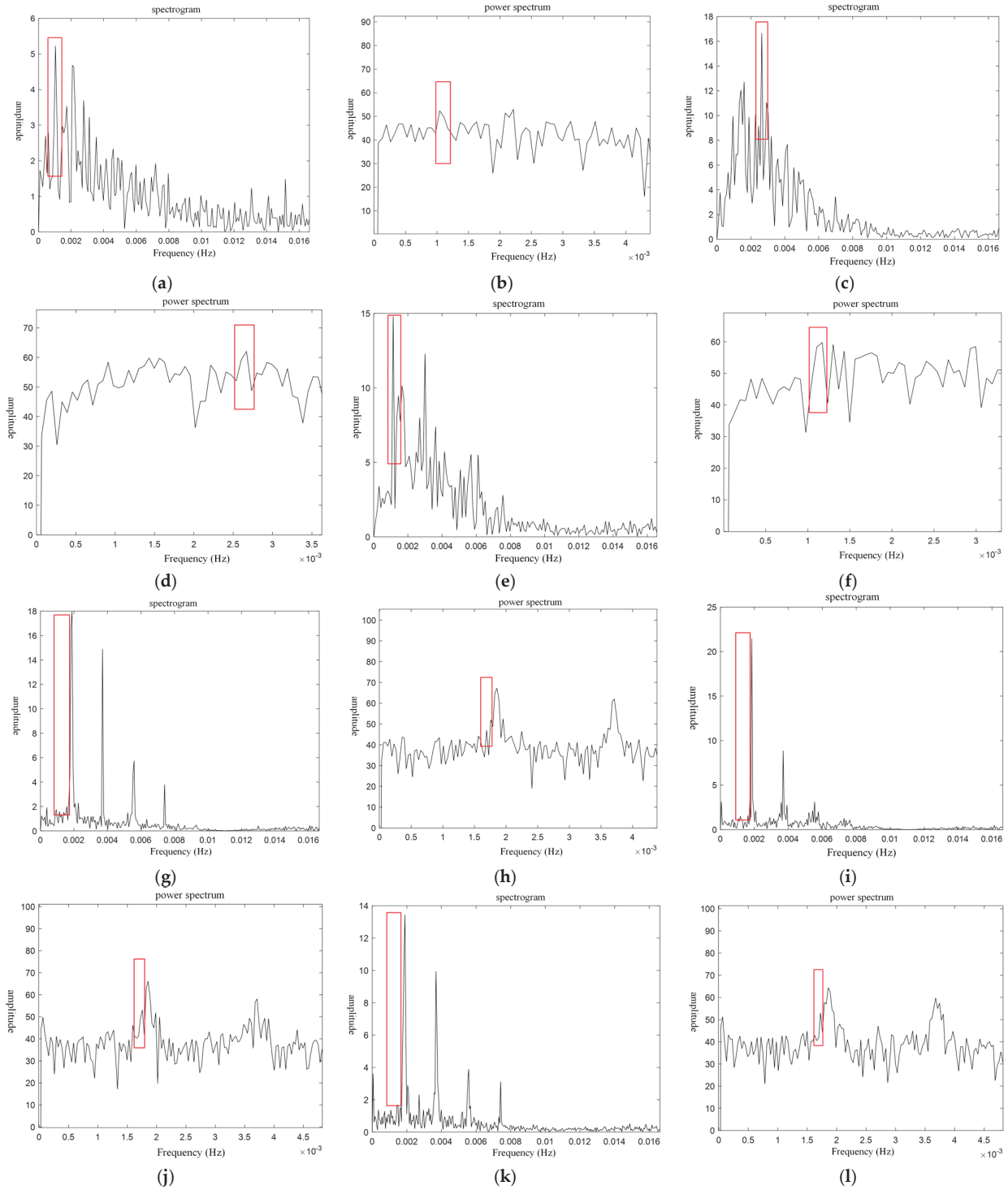


Figure 8. Survey roadway flow map. (**a**) Flow spectrogram of section A. (**b**) Flow power spectra of section A. (**c**) Flow spectrogram of section B. (**d**) Flow power spectra of section B. (**e**) Flow spectrogram of section C. (**f**) Flow power spectra of section C. (**g**) Flow spectrogram of section D. (**h**) Flow power spectra of section D. (**i**) Flow spectrogram of section E. (**j**) Flow power spectra of section E. (**k**) Flow spectrogram of section F. (**l**) Flow power spectra of section F.

In the power spectrum model, the main peak represents the highest point in the graph, and its intensity reflects the maximum stable flow rate for the road. In this study, the intensity of the main peak in the power spectrum is converted to non-motorized road flow per unit time and unit width, representing the non-motorized road capacity targeted in this analysis.

From the power spectrum image, the primary eigenvalues were extracted, as shown in Table 9.

Table 9. Flow power spectrum feature extraction.

Sections	Main Valve Strength	Frequency Values
Section A	52.3595	0.001042
Section B	62.0679	0.002669
Section C	59.8066	0.001172
Section D	62.2261	0.001855
Section E	64.1054	0.001855
Section F	62.3808	0.001855

To calculate capacity, traffic flow is uniformly converted to the flow of mixed non-motorized vehicles per unit time and unit width, expressed in bikes/(h·m). The calculation formula is provided in Equation (24).

$$Q = \frac{q}{n \times w} \times 3600 = \frac{q}{n \times (W - 0.5)} \times 3600 \quad (24)$$

In the equation, q represents the number of mixed non-motorized vehicles passing within the statistical interval; q is the primary lobe intensity value obtained from the power spectrum model; n refers to the time interval, which in this paper is set to 30 s, w represents the effective width of the non-motorized vehicle lane; and W refers to the total width of the non-motorized vehicle lane.

The traffic flows examined in this study underwent vehicle equivalent calculations within the stochastic capacity estimation model based on power spectrum analysis. Accordingly, the capacity estimates for the selected road sections were converted to equivalent pure EV and bicycle capacities, as shown in Table 10.

Table 10. Estimated capacity.

Sections	Pure Electric Vehicle Capacity (Bikes/(h·m))	Pure Bicycle Capacity (Bikes/(h·m))
Section A	2060	1538
Section B	2979	2223
Section C	3262	2434
Section D	2370	1769
Section E	3077	2296
Section F	3297	2460

To validate the effectiveness of the proposed traffic capacity estimation model, a comparison was made with existing density–flow relationship models. Figure 9 shows the scatter plots and fitting curves for the density–flow relationship in non-motorized lanes of sections D, E, and F. Observing the scatter plots reveals that the overall trend of the points follows a power function form. After comparing several power functions, the quadratic function was found to provide the best fit, so it was selected to model the flow-density distribution. The fitted density–flow models and calculation results for each section are presented in Table 11.

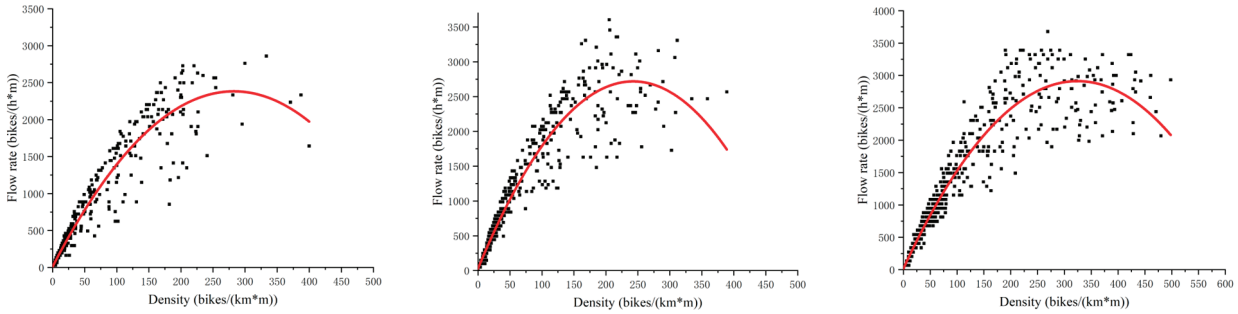


Figure 9. Density–flow map of surveyed roadway segments. (**a**) Density–flow diagram for section D. (**b**) Density–flow diagram for section E. (**c**) Density–flow diagram for section F.

Table 11. Density–flow model fit table.

Sections	(Math.) Simultaneous Equations	R^2	Pure Electric Vehicle Capacity (Bikes/(h·m))	Pure Bicycle Capacity (Bikes/(h·m))
Section A	$Q = -0.0279k^2 + 16.2446k$	0.928	2349	1753
Section B	$Q = -0.04485k^2 + 21.9347k$	0.938	2647	1975
Section C	$Q = -0.02572k^2 + 16.95613k$	0.941	2867	2139
Section D	$Q = -0.0297k^2 + 16.78492k$	0.933	2382	1778
Section E	$Q = -0.04568k^2 + 22.17318k$	0.932	2718	2028
Section F	$Q = -0.02752k^2 + 17.85613k$	0.934	2912	2173

The comparison shows that the stochastic capacity model developed in this study aligns closely with the results of the density–flow model. This validates the stochastic capacity model, further enriches the theory of mixed non-motorized traffic flow, and provides theoretical support for non-motorized lane management.

5. Conclusions

This study focuses on mixed non-motorized vehicles on urban road sections, investigating traffic flow characteristics and capacity in non-motorized lanes. The main study conclusions are as follows:

(1) UAVs and cameras were used to collect video data on non-motorized lanes with hard segregation on urban roads, capturing mixed non-motorized traffic flow characteristics. Preliminary analyses were conducted on rider and vehicle characteristics. Speed data were fitted to normal, lognormal, gamma, and Weibull distributions, demonstrating the stochasticness and volatility of mixed traffic flow.

(2) A stochastic capacity model was constructed based on power spectrum estimation. By analyzing stochastic signal characteristics, this study introduces the frequency domain concept to investigate changes in traffic flow characteristics. Using power spectrum estimation, the stochastic signal model is fitted to traffic flow data, with AR, MA, ARMA, and ARIMA models selected based on autocorrelation and partial autocorrelation functions. Optimal parameters are chosen via the AIC criterion, followed by time–frequency domain transformation. The primary peak intensity in the power spectrum is used to measure passability.

(3) An empirical validation of the stochastic capacity model was conducted. The calculations of the investigated traffic flow data estimate pure electric vehicle (EV) capacity at 2060–3297 bikes/(h·m) and pure bicycle capacity at 1538–2460 bikes/(h·m). The density–flow model estimates place capacity at 2349–2897 bikes/(h·m) for pure electric vehicles and 1753–2173 bikes/(h·m) for pure bicycles. The comparison confirms that the capacity estimates from the stochastic capacity model align with existing studies.

Author Contributions: Conceptualization, G.G., X.S., B.L. and J.C.; methodology, G.G., X.S., B.L., B.Y. and X.W.; validation, X.S., B.L., B.Y., D.Z. and S.H.; formal analysis, B.L., X.W., Q.H. and C.B.; investigation, X.S., S.H. and Q.H.; resources, G.G. and J.C.; data curation, X.S., B.L., and B.Y.; writing—original draft preparation, X.S., B.L., B.Y. and X.W.; writing—review and editing, B.Y., X.W., J.C. and D.Z.; visualization, B.L., S.H. and Q.H.; supervision, G.G., J.C. and D.Z.; project administration, G.G. and J.C.; funding acquisition, G.G., J.C. and C.B. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by Guangxi Science and Technology Base and Talent Project for Guangxi Science and Technology Plan Project: Construction of Guangxi Transportation New Technology Transfer Center Platform (No. Guike AD23026029) and Construction of Scientific and Technological Innovation Bases for Science Research and Technology Development Plan Project of Nanning City: Intelligent Operation and Maintenance of Transportation Infrastructure and Health Monitoring Platform (No. 20223230). Transportation Operation Subsidy Project of Guangxi Key Laboratory of International Join for China-ASEAN Comprehensive Transportation in 2021 (No. 21-220-21).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets presented in this article are not readily available because the data are part of an ongoing study. Requests to access the datasets should be directed to chenjianqiu@unn.edu.cn.

Conflicts of Interest: Author Xin Sun was employed by the company China Nuclear Engineering Consulting Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Böther, M.; Schiller, L.; Fischbeck, P.; Molitor, L.; Krejca, M.S.; Friedrich, T. Evolutionary minimization of traffic congestion. *IEEE Trans. Evol. Comput.* **2023**, *27*, 1809–1821. [CrossRef]
2. Shen, T.; Hong, Y.; Thompson, M.M.; Liu, J.; Huo, X.; Lian, W. How does parking availability interplay with the land use and affect traffic congestion in urban areas? The case study of Xi'an, China. *Sustain. Cities Soc.* **2020**, *57*, 102126. [CrossRef]
3. Jiang, J.; Tu, W.; Kong, H.; Zeng, W.; Zhang, R.; Konečný, M. Large-scale urban multiple-modal transport evacuation model for mass gathering events considering pedestrian and public transit system. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 23059–23069. [CrossRef]
4. Andor, M.; Gerster, A.; Gillingham, K.; Horvath, M. Running a car costs much more than people think—Stalling the uptake of green travel. *Nature* **2020**, *580*, 453–455. [CrossRef]
5. Geng, J.; Long, R.; Chen, H.; Li, W. Exploring the motivation-behavior gap in urban residents' green travel behavior: A theoretical and empirical study. *Resour. Conserv. Recycl.* **2017**, *125*, 282–292. [CrossRef]
6. Patella, S.M.; Grazieschi, G.; Gatta, V.; Marcucci, E.; Carrese, S. The adoption of green vehicles in last mile logistics: A systematic review. *Sustainability* **2020**, *13*, 6. [CrossRef]
7. Stefánsdóttir, H.; Næss, P.; Ihlebæk, C. Built environment, non-motorized travel and overall physical activity. *Travel Behav. Soc.* **2019**, *16*, 201–213. [CrossRef]
8. Montanino, M.; Punzo, V. On string stability of a mixed and heterogeneous traffic flow: A unifying modelling framework. *Transp. Res. Part B Methodol* **2021**, *144*, 133–154. [CrossRef]
9. Liu, X.; Chen, Q. An algorithm for the mixed transportation network design problem. *PLoS ONE* **2016**, *11*, e0162618. [CrossRef]
10. He, P.; Pan, W. Collaborative optimization model and algorithm for airport capacity and traffic flow allocation. *PLoS ONE* **2024**, *19*, e0298540. [CrossRef]
11. Qin, J.; Ma, S.; Zhang, L.; Wang, Q.; Feng, G. Modeling and simulation for non-motorized vehicle flow on road based on modified social force model. *Mathematics* **2022**, *11*, 170. [CrossRef]
12. Hu, S.; Tong, W.; Jia, Z.; Zou, J. Study on the spatial and temporal distribution and traffic flow parameters of non-motorized vehicles on highway segments crossing small towns. *Sustainability* **2023**, *15*, 1261. [CrossRef]
13. Chen, J. Operating speed of mixed traffic flow of motorized vehicle and non-motorized vehicle on urban road. *Comput. Commun.* **2008**, *26*, 14–17.
14. Bigazzi, A.; Lindsey, R. A utility-based bicycle speed choice model with time and energy factors. *Transportation* **2018**, *46*, 995–1009. [CrossRef]
15. Kalair, K.; Connaughton, C. Anomaly detection and classification in traffic flow data from fluctuations in the flow-density relationship. *Transp. Res. Part C Emerg. Technol.* **2021**, *127*, 103178. [CrossRef]

16. Jin, S.; Xu, L.; Cheng, X.; Ma, D. Lane width-based cellular automata model for mixed bicycle traffic flow. *Comput.-Aided Civ. Infrastruct. Eng.* **2019**, *34*, 696–712. [CrossRef]
17. Grigoropoulos, G.; Leonhardt, A.; Kaths, H.; Junghans, M.; Baier, M.; Busch, F. Traffic flow at signalized intersections with large volumes of bicycle traffic. *Transp. Res. Part A Policy Pract.* **2022**, *155*, 464–483. [CrossRef]
18. Wierbos, M.J.; Knoop, V.L.; Hänseler, F.; Hoogendoorn, S.P. A macroscopic flow model for mixed bicycle–car traffic. *Transportmetrica* **2020**, *17*, 340–355. [CrossRef]
19. Feng, T.; Liu, J.-Y.; Liang, C.; Tian, X.; Chen, C.; Liu, K. Analysis of left-turn behaviors of non-motorized vehicles and vehicle-bicycle conflicts. *PLoS ONE* **2023**, *18*, e0291504. [CrossRef]
20. Liu, Y.; Alsaleh, R.; Sayed, T. Modelling motorized and non-motorized vehicle conflicts using multiagent inverse reinforcement learning approach. *Transportmetr. B Transp. Dyn.* **2024**, *12*, 2314762. [CrossRef]
21. Gu, J.; Jia, Z.; Cai, T.; Song, X.; Mahmood, A. Dynamic Correlation Adjacency-Matrix-Based Graph Neural Networks for Traffic Flow Prediction. *Sensors* **2023**, *23*, 2897. [CrossRef] [PubMed]
22. Li, G.; Deng, H.; Yang, H. A Multi-Factor Combined Traffic Flow Prediction Model with Secondary Decomposition and Improved Entropy Weight Method. *Expert Syst. Appl.* **2024**, *255*, 124424. [CrossRef]
23. Huan, W.; Po, H.; Hao, W. A Genetic Timing Scheduling Model for Urban Traffic Signal Control. *Inf. Sci.* **2021**, *576*, 475–483.
24. Wen, H.; Huan, W.; Liping, Y.; Bo, D. A Swarm Intelligent Method for Traffic Light Scheduling: Application to Real Urban Traffic Networks. *Appl. Intell.* **2016**, *44*, 208–231.
25. Çelebi, D.; Yörüsün, A.; Işık, H. Bicycle sharing system design with capacity allocations. *Transp. Res. Part B Methodol.* **2018**, *114*, 86–98. [CrossRef]
26. Navin, F. Bicycle traffic flow characteristics: Experimental results and comparisons. *ITE J. Inst. Transp. Eng.* **1994**, *64*, 31–37.
27. Botma, H. Method to determine level of service for bicycle paths and pedestrian-bicycle paths. *Transp. Res. Rec.* **1995**, *1502*, 38–44.
28. Liu, X.; Liu, C.; Ren, F. Operational analysis of bicycle interchanges in Beijing, China. *Transp. Res. Rec.* **1993**, *1396*, 18–21.
29. Guo, N.; Wong, W.H.; Jiang, R.; Wong, S.C.; Hao, Q.; Wu, C.Y. Bicycle flow dynamics of cyclist loading and unloading processes at bottlenecks. *Transp. Sci.* **2023**, *58*, 340–354. [CrossRef]
30. Brown, M.; Scott, D.M.; Páez, A. A spatial modeling approach to estimating bike share traffic volume from GPS data. *Sustain. Cities Soc.* **2022**, *76*, 103401. [CrossRef]
31. Giuffrè, O.; Granà, A.; Tumminello, M.L.; Sferlazza, A. Estimation of passenger car equivalents for single-lane roundabouts using a microsimulation-based procedure. *Expert Syst. Appl.* **2017**, *79*, 333–347. [CrossRef]
32. Sugiarto, S.; Apriandy, F.; Darma, Y.; Saleh, S.M.; Rusdi, M.; Miwa, T. Determining passenger car equivalent (PCEs) for pre-timed signalized intersections with severe motorcycle composition using Bayesian linear regression. *PLoS ONE* **2021**, *16*, e0256620. [CrossRef]
33. Yeung, J.S.; Wong, Y.D.; Secadiningrat, J.R. Lane-harmonised passenger car equivalents for heterogeneous expressway traffic. *Transp. Res. Part A Policy Pract.* **2015**, *78*, 361–370. [CrossRef]
34. Giuffrè, O.; Granà, A.; Tumminello, M.L.; Sferlazza, A. Capacity-based calculation of passenger car equivalents using traffic simulation at double-lane roundabouts. *Simul. Model. Pract. Theory* **2018**, *81*, 11–30. [CrossRef]
35. Zhang, S.; Tak, T.-O. Efficiency evaluation of electric bicycle power transmission systems. *Sustainability* **2021**, *13*, 10988. [CrossRef]
36. Wang, D.; Chen, L.; Jing, C.; Wang, G. Bicycle conversion factor calibration at two-phase intersections in mixed traffic flows. *Tsinghua Sci. Technol.* **2007**, *12*, 318–323. [CrossRef]
37. Schaefer, J.S.; Figliozzi, M.; Unnikrishnan, A. Evidence from urban roads without bicycle lanes on the impact of bicycle traffic on passenger car travel speeds. *Transp. Res. Rec.* **2020**, *2674*, 87–98. [CrossRef]
38. Li, Y.; Yuan, H.; Wang, Y.; Xiao, C. GGT-YOLO: A novel object detection algorithm for drone-based maritime cruising. *Drones* **2022**, *6*, 335. [CrossRef]
39. Pacal, I.; Karaman, A.; Karaboğa, D.; Akay, B.; Baştürk, A.; Nalbantoğlu, U.; Coskun, S. An efficient real-time colonic polyp detection with YOLO algorithms trained by using negative samples and large datasets. *Comput. Biol. Med.* **2022**, *141*, 105031. [CrossRef]
40. Feng, K.; Huo, W.; Xu, W.; Li, M.; Li, T. CNA-DeepSORT algorithm for multi-target tracking. *Multimed. Tools Appl.* **2023**, *83*, 4731–4755. [CrossRef]
41. Huang, K.; Chu, J.; Leng, L.; Dong, X. TATrack: Target-aware transformer for object tracking. *Eng. Appl. Artif. Intell.* **2024**, *127*, 107304. [CrossRef]
42. Xiong, X.; He, M.; Li, T.; Zheng, G.; Xu, W.; Fan, X.; Zhang, Y. Adaptive feature fusion and improved attention mechanism-based small object detection for UAV target tracking. *IEEE Internet Things J.* **2024**, *11*, 21239–21249. [CrossRef]
43. Turvey, G.J.; Milburn, J. Two-bay pultruded GFRP safety barrier/guardrail—Testing, analysis and compliance with standards. *Compos. Struct.* **2020**, *253*, 112745. [CrossRef]
44. Fu, L.; Zhang, Y.; Qin, H.; Shi, Q.; Chen, Q.; Shi, Y. A modified social force model for studying nonlinear dynamics of pedestrian-e-bike mixed flow at a signalized crosswalk. *Chaos Solitons Fractals* **2023**, *174*, 113813. [CrossRef]
45. Liu, Y.; Fu, C.; Wang, W. Modeling duration of overtaking between non-motorized vehicles: A nonparametric survival analysis-based approach. *PLoS ONE* **2021**, *16*, e0244883. [CrossRef]

46. Maris, E. A bicycle can be balanced by stochastic optimal feedback control but only with accurate speed estimates. *PLoS ONE* **2023**, *18*, e0278961. [CrossRef]
47. Wang, Z.; Liu, Z.; Cheng, Q.; Gu, Z. Integrated self-consistent macro-micro traffic flow modeling and calibration framework based on trajectory data. *Transp. Res. Part C Emerg. Technol.* **2024**, *158*, 104439. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

applied sciences

MDPI

Article

LG-YOLOv8: A Lightweight Safety Helmet Detection Algorithm Combined with Feature Enhancement

Zhipeng Fan [1,2,*], Yayun Wu [1], Wei Liu [1,2], Ming Chen [1,2] and Zeguo Qiu [1,2]

[1] School of Computer and Information Engineering, Harbin University of Commerce, Harbin 150028, China; wuyayun@s.hrbcu.edu.cn (Y.W.); weiwei@hrbcu.edu.cn (W.L.); chenming@hrbcu.edu.cn (M.C.); qiuze@hrbcu.edu.cn (Z.Q.)

[2] Heilongjiang Provincial Key Laboratory of Electronic Commerce and Information Processing, Harbin 150028, China

* Correspondence: fanzhipeng@hrbcu.edu.cn

Abstract: In the realm of construction site monitoring, ensuring the proper use of safety helmets is crucial. Addressing the issues of high parameter values and sluggish detection speed in current safety helmet detection algorithms, a feature-enhanced lightweight algorithm, LG-YOLOv8, was introduced. Firstly, we introduce C2f-GhostDynamicConv as a powerful tool. This module enhances feature extraction to represent safety helmet wearing features, aiming to improve the efficiency of computing resource utilization. Secondly, the Bi-directional Feature Pyramid (BiFPN) was employed to further enrich the feature information, integrating feature maps from various levels to achieve more comprehensive semantic information. Finally, to enhance the training speed of the model and achieve a more lightweight outcome, we introduce a novel lightweight asymmetric detection head (LADH-Head) to optimize the original YOLOv8-n's detection head. Evaluations on the SWHD dataset confirm the effectiveness of the LG-YOLOv8 algorithm. Compared to the original YOLOv8-n algorithm, our approach achieves a mean Average Precision (mAP) of 94.1%, a 59.8% reduction in parameters, a 54.3% decrease in FLOPs, a 44.2% increase in FPS, and a 2.7 MB compression of the model size. Therefore, LG-YOLOv8 has high accuracy and fast detection speed for safety helmet detection, which realizes real-time accurate detection of safety helmets and an ideal lightweight effect.

Keywords: safety helmet testing; lightweight network; C2f-GhostDynamicConv; BiFPN; LADH-Head

1. Introduction

As industrialization advances, the emphasis on safety management in the construction industry has grown. Safety helmets, which are essential personal protective equipment, play a vital role on construction sites, substantially reducing the potential for head injuries. However, the consistent and conscientious use of these helmets by workers is not always guaranteed, highlighting the need for real-time monitoring to ensure their utilization and, by extension, secure work environments. Unfortunately, such monitoring is typically resource-intensive and laborious, and it also poses challenges in terms of cost and practicality when it comes to deploying on edge devices for an efficient outcome.

Currently, certain accomplishments have been attained in the domain of object detection. Particularly, the approaches based on deep learning have made breakthrough advancements in the performance detection of numerous tasks. Presently, the mainstream methods are mainly categorized into two types: two-stage algorithms and single-stage algorithms. The former mainly relies on the two-stage object detection algorithm of candidate regions to achieve high-precision object detection. Representative algorithms in this category include R-CNN [1], Faster R-CNN [2], etc. The second category uses a single network to predict the location and category of the object directly from the input image, avoiding the complex process of generating candidate regions and achieving faster detection speed, and the main representative networks in this category include SSD [3] and YOLO [4–9].

These two types of methods have certain advantages for target detection tasks, but considering the specific task of safety helmet detection and in the real scene of a construction site for objects of different sizes and the influence of other uncontrollable factors, especially for some places with limited resources, it is a general trend to design high precision lightweight models. YOLOv8-n [10] demonstrates enhanced accuracy and speed over its predecessors, yet the detection accuracy for small objects like helmets remains suboptimal, and the substantial model size complicates deployment on edge devices. Based on this, on the basis of the YOLOv8-n algorithm, this paper optimizes the model to a certain extent and introduces the novel LG-YOLOv8 safety helmet detection algorithm, which is a high-precision lightweight model with fast detection speed and strong detection effect.

The main contributions of this paper are as follows:

(1) In the backbone of the YOLOv8-n model, a novel structure named C2f-Ghost-DynamicConv is introduced. This structure amalgamates the strengths of the Ghost module and dynamic convolution, substantially decreasing computational load and parameter count through a distinctive feature generation process. Concurrently, it employs dynamic convolution to adaptively modify the convolution kernel, thereby enhancing the model's feature representation capabilities. This innovation notably improves the detection performance in complex construction settings, such as those involving distant or partially obscured helmets, leading to heightened accuracy and efficiency.

(2) Optimize the traditional unidirectional feature fusion method into a bidirectional feature pyramid structure, enhancing feature fusion and multi-scale feature extraction capabilities through the bidirectional feature pyramid network. By leveraging contextual information and combining top-down and bottom-up feature transfer, it effectively extracts multi-scale features, addressing the issues of background interference and leakage in complex scenes and different sizes of helmets. This improves the model's recognition accuracy for helmets.

(3) In the construction of the detection head, the LADH-Head has been implemented as an upgrade. It utilizes local dense connections and an adaptive module to significantly improve the detailed features of small targets such as safety helmets, thus minimizing information loss and substantially reducing the occurrence of missed and false detections of safety helmets. Furthermore, the application of the asymmetric convolution operation serves to diminish the parameter count and computational demand. This not only lightens the hardware resource demands of the model but also ensures exceptional detection performance.

(4) The proposed lightweight detection algorithm network was evaluated on the SHWD dataset, and the mAP and FPS reached 94.1% and 153.3, respectively, and the GFLOPs also reached 3.7. Compared with other methods, the accuracy of the model is improved, the execution speed is also significantly improved, and the computational cost is significantly reduced.

2. Related Work

The application of deep learning in the field of helmet detection has gradually become a research hotspot, and some remarkable research results have been obtained. Lian et al. [11] introduced a novel network, HR-YOLO, featuring a residual network architecture and a PANMS method to enhance safety helmet detection precision. Han et al. [12] designed a lightweight end-to-end convolutional neural network, GhostNet, as the backbone, incorporating multi-scale segmentation and feature fusion during feature processing. Using this design, the accuracy reached 93.5%, and the FPS also achieved good results. Chen et al. [13] utilized the K-means++ algorithm to categorize helmet sizes in images and subsequently applied an enhanced Faster-RCNN algorithm for helmet usage detection. These studies predominantly concentrate on elevating detection accuracy and efficiency, yet this often escalates model complexity, computation demands, and parameter counts, overlooking the constrained computational capabilities of embedded systems at construction sites. The advent of edge computing has seen an increasing number of surveillance

systems implementing recognition algorithms on edge-embedded devices, catalyzing the emergence of lightweight target detection algorithms as a burgeoning research focus. Chen et al. [14] employed PP-LCNet as the backbone of YOLOv4 and incorporated depthwise separable convolution to decrease model parameters, thereby enhancing both accuracy and detection speed.

Based on the above research, we conclude that the current focus is on improving the accuracy and efficiency of object detection, but this leads to increasing model complexity, increasing the amount of calculation and parameters, so lightweight object detection algorithms have gradually become a new research hotspot, and many lightweight networks have emerged. Examples are ShuffleNetV2 [15], MobileNetV2 [16], GhostNet [17], etc. As a result, we extensively optimize the architecture of YOLOv8-n. Our goal is not only to improve the detection accuracy but also to reduce the computational overhead and improve the operating efficiency.

3. Materials and Methods

3.1. The Original YOLOv8 Algorithm

This paper builds upon YOLOv8-n version 8.1, introduced by the Ultralytics team in 2023. YOLOv8-n ensures smooth interoperability with prior YOLO iterations, enabling direct performance evaluations. The YOLOv8-n algorithm, incorporating fresh enhancements over YOLOv5, exhibits remarkable performance in object detection, achieving a previously unmatched equilibrium between speed and accuracy [18,19]. The algorithm's architecture is primarily segmented into input, backbone network, neck network, and output, as depicted in Figure 1.

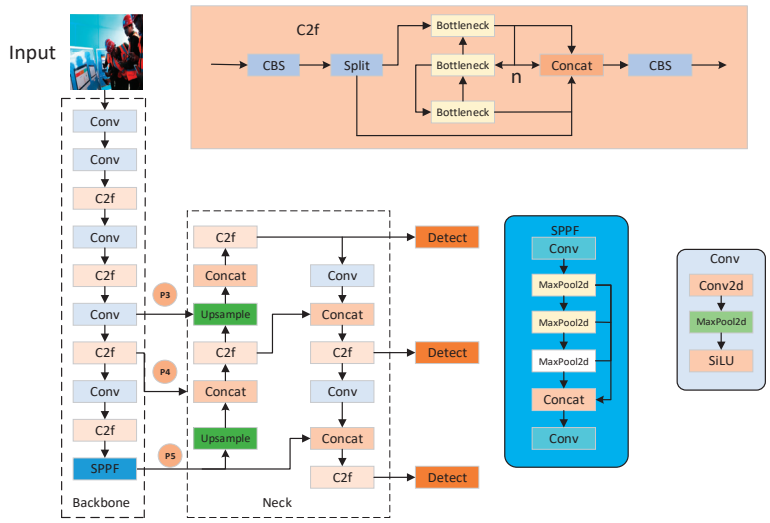


Figure 1. Network structure of YOLOv8.

The backbone network is primarily tasked with feature extraction. The YOLOv8-n algorithm replaces the earlier C3 structure [20] with the C2f structure, which enriches gradient flow and tailors the channel count to suit different algorithm scales. The neck network is comprised of two components: the Feature Pyramid Network (FPN) [21] and the Path Aggregation Network (PAN) [22], which together construct the feature pyramid. PANet merges bottom-up and top-down strategies to effectively capture multi-scale information, aiding in the fusion of information across various feature layers. This enhancement allows the detector to more adeptly accommodate objects of diverse sizes and shapes, thereby boosting detection accuracy. Furthermore, YOLOv8-n omits the 1×1 convolution prior to upsampling within the neck component, thereby diminishing computational load. In the head component, the conventional decoupled-head structure [23] is employed to segregate the classification and detection functionalities, aiming to mitigate the inherent tension between classification and localization objectives. YOLOv8-n applies the BCE loss for

classification tasks, while the regression loss integrates the focus loss, specifically the DFL loss, along with the CIOU loss.

3.2. LG-YOLOv8 Algorithm

Enhancing the efficiency of the YOLOv8-n algorithm for detecting safety helmets by optimizing its lightweight capabilities, this study introduces the LG-YOLOv8. The algorithm is characterized by optimizing the original C2f module by introducing the C2f-GhostDynamicConv module, an advanced dynamic convolution method. This module enhances feature representation by dynamically adjusting the convolution kernel, allowing the model to more accurately capture details and adapt to diverse detection contexts. Simultaneously, this innovation significantly reduces the model's parameter count and computational load, providing a robust foundation for the deployment of lightweight object detection. Subsequently, the unidirectional pyramid network in the neck section is upgraded to a bidirectional pyramid network (BiFPN), a configuration that enriches the representation of feature maps across different scales via a bidirectional fusion approach, thereby further elevating detection precision. Lastly, the detection head is innovatively replaced with LADH-Head, a lightweight asymmetric detection head, which ensures accurate detection of targets of varying sizes while preserving the model's operational efficiency through the asymmetric multi-stage compression technique and dual-head architecture. By integrating these modules, the YOLOv8-n model not only retains its computational efficiency advantage but also achieves a significant leap in target detection accuracy and adaptability, particularly in handling complex visual environments, rendering it more suitable for deployment on devices with constrained computational resources.

3.2.1. GhostDynamicConv

To facilitate the model's operation on edge devices for accurate helmet detection within smart construction site settings that have constrained computational resources, we have integrated the C2f-GhostDynamicconv module to refine the detection model. This architecture is designed to augment the C2f component of the backbone network by harnessing the combined advantages of the Ghost module and dynamic convolution, utilizing a novel feature extraction method. It selectively merges various convolution kernels in response to the incoming data and adapts the weights, thus promoting the identification and portrayal of features pertinent to helmets. This strategy adeptly tackles the difficulties associated with identifying small objects, dense crowds, and intricate backgrounds within helmet detection contexts. Moreover, this module optimizes the usage of computational resources and enhances the efficacy of operations performed with reduced precision floating-point arithmetic.

- GhostModule

The traditional convolution method has a large amount of calculation, which cannot eliminate redundant features well for safety helmet detection, and the smaller convolution kernel cannot reduce memory and FLOP to a certain extent. Han et al. [17] introduced the Ghost module, with the core idea of GhostNet being to produce additional feature maps through inexpensive operations and fewer parameters. The output of regular convolution may have redundancy, and this kind of redundant feature map can be obtained by simple linear operation based on a certain base feature. The Ghost module, as the base module of GhostNet, extracts rich feature information by regular convolution and generates the redundant feature information by adopting inexpensive linear transformation, which can significantly reduce the computational resources needed by the model. Figure 2 shows the schematic diagram of conventional convolution and Ghost convolution, both of which can be specifically expressed by the formula, where x denotes the cheap linear operation, and both of them will output the same number of feature maps.

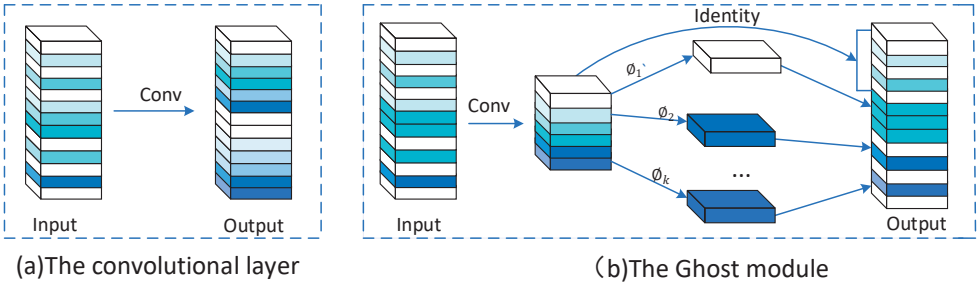


Figure 2. Conventional convolution and Ghost module. (**a**) The convolutional layer; (**b**) The Ghost module.

Standard convolution:

$$Y = \mathrm{Conv}(X, W) \quad \text{where} \quad Y \in R^{H' \times W' \times C'} \tag{1}$$

Cheap transformations:

$$Y^{ghost} = \{\phi_1(Y), \phi_2(Y), \cdots, \phi_k(Y)\} \quad \text{where} \quad k = \frac{s \times C'}{C'} = s \tag{2}$$

In this scenario, if the input feature map is of size $H \times W \times C$, the standard convolution produces a basic feature map of size $H' \times W' \times C'$, where C' is typically less than C. Subsequently, the "artifacts" are generated through cost-effective transformation s times. This results in s-fold "artifacts" produced by the inexpensive transformation, leading to a total of $s \times C'$ feature maps. In the design principle of GhostModule, depthwise separable convolutions are employed to carry out efficient linear transformations, thereby generating the so-called "Ghost feature maps". These feature maps are produced by depthwise separable convolution, where s represents the standard convolution operation and ϕ indicates the linear transformation implemented through depthwise separable convolution. With this design, GhostModule can considerably reduce the number of parameters and computational cost while preserving the accuracy of the network.

Assuming that there are s redundant features for each base feature, then only n/s base features need to be output, followed by an inexpensive linear transformation of each base feature to generate the corresponding redundant features. This reduces the computation of the model, and the GFLOPS is 1/s of the standard convolution. Let the average size of the convolution kernel used in the linear operation part be $d \times d$, the theoretical speedup ratio r, as shown in Equation (3). In the context of the output feature layer, h denotes the length and w represents the width. The variable k signifies the dimension of the standard convolution kernel, while n indicates the total number of convolutions performed. The parameter c corresponds to the channel count of the input. If d squared is equivalent to k squared and s is no greater than c, the theoretical increase in speed is s times that of the initial configuration:

$$r = \frac{n \cdot h \cdot w \cdot c \cdot k \cdot k}{\frac{n}{s} \cdot h \cdot w \cdot c \cdot k \cdot k + (s-1)\frac{n}{s} \cdot h \cdot w \cdot d \cdot d} = \frac{c \cdot k \cdot k}{\frac{1}{s} \cdot c \cdot k \cdot k + \frac{s-1}{s} \cdot d \cdot d} \approx \frac{s \cdot c}{s + c - 1} \approx s \tag{3}$$

- C2f-GhostDynamicConv

Large-scale visual training tasks such as helmet detection often require a low FLOP model that can still provide good detection advantages; therefore, DynamicConv was merged with GhostBottleneck and incorporated into the initial C2f module, yielding the C2f-GhostDynaimicConv module. Dynamic convolution, as discussed in the literature [24,25], improves feature representation over traditional static convolution. The main core idea is to dynamically generate convolution kernels based on input features or to fine-tune the convolution kernels. Dynamic convolution can adaptively adjust the weights of the convolution kernel according to the characteristics of the input image, which can better capture the detailed features of the helmet, in addition to dealing with environmental

changes, angular changes, and occlusion problems, and improve the robustness of the detection rate in a variety of site environments. DynamicConv is based on a dynamic perception whose basic principle is illustrated in the following equations:

$$y = g\left(\tilde{W}^{T}(x)x + \tilde{b}(x)\right) \tag{4}$$

$$\tilde{W}(x) = \sum_{k=1}^{k} \pi_k \tilde{W}_k, \tilde{b}(x) = \sum_{k=1}^{K} \pi_k(x)\tilde{b}_k \tag{5}$$

$$\text{s.t. } 0 \leq \pi_k(x) \leq 1, \sum_{k=1}^{k} \pi_k(x) = 1 \tag{6}$$

Within the given formula, x signifies the input data, while y represents the resulting output. It is evident that x is subjected to a pair of separate processes: initially, it serves as a parameter for the attention mechanism that formulates the convolution kernel, and subsequently, it is applied in the convolution process itself. The variables W, b, and g correspond to the weight matrix, bias vector, and activation function, respectively. The term $\pi_k(x)$ indicates the attention coefficient for the k-th linear transformation, highlighting that this coefficient is contingent upon the input x. The kernel is composed of a convolutional filter intertwined with the k-th linear transformation. The dynamic convolution is fundamentally made up of k convolutional kernels and adheres to the conventional architecture of convolutional neural networks (CNNs). It incorporates batch normalization (BatchNorm) and the Rectified Linear Unit (ReLU) as its activation function. Figure 3 provides a detailed depiction of the architecture of DynamicConv.

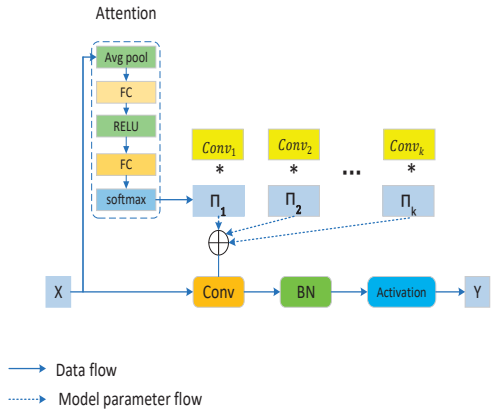


Figure 3. The structure of DynamicConv.

In contrast to DynamicConv, which is predominantly concerned with channel-wise attention, SENet, or Squeeze-and-Excitation Networks [26], treats the entire convolutional kernel as a focus of attention. By extending the concepts of DynamicConv and adopting the architectural innovations of the GhostBottleneck module, the initial C2f module was upgraded to form the C2f-GhostDynamicConv module. This innovative design markedly enhances precision with a minimal increase in the computational complexity, measured in FLOPs. Figure 4 contrasts this innovative design with the C2f module in the baseline model YOLOv8-n.

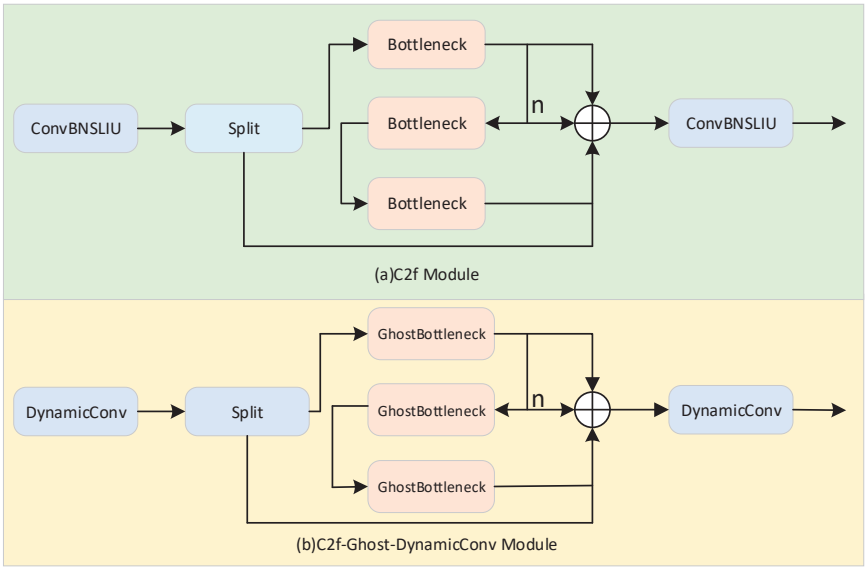


Figure 4. Comparison of C2f and C2f-GhostDynamicConv modules: (**a**) C2f module; (**b**) C2f-GhostDynamicConv module.

3.2.2. BiFPN

The neck design of YOLOv8 draws inspiration from PANet [27], enhancing feature fusion by combining the Path Aggregation Network (PAN) and Feature Pyramid Network (FPN) architectures, as illustrated in Figure 5a,b. Unlike YOLOv5 [28] and YOLOv7, YOLOv8's FPN architecture employs top-down fusion, where high-level features are upsampled and integrated into low-level features, enriching them with semantic details. Conversely, PAN adopts a bottom-up strategy, downsampling low-level features and incorporating them into high-level features, thereby augmenting spatial precision. The integration of these techniques forms the PANet structure, facilitating adaptive fusion of shallow and deep features, which substantially enhances the model's capability to discern fine details. Nonetheless, when merging features of disparate resolutions, the features are typically aggregated through straightforward summation, resulting in unequal contributions from various layers. As the network progresses deeper, the impact of high-level features intensifies, while low-level features diminish. To address this, TAN et al. [29] proposed BiFPN to produce feature maps with more enriched semantic content.

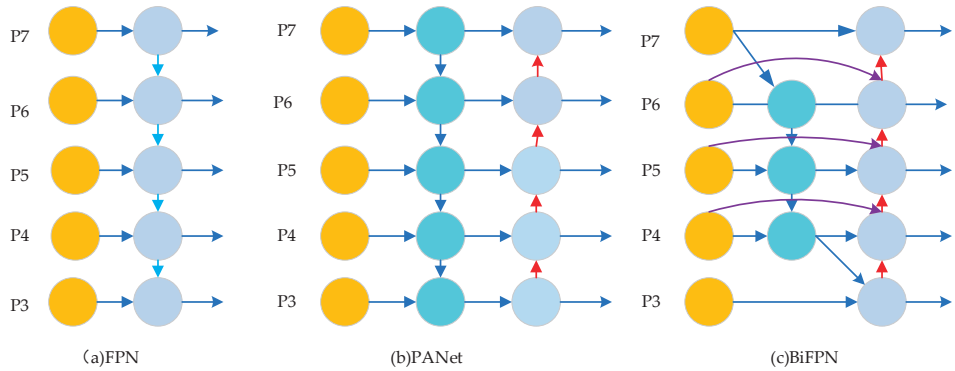


Figure 5. (**a**) FPN introduces a top-down path to fuse multiscale features from the third to the seventh level (P3–P7); (**b**) PANet enhances the FPN (Feature Pyramid Network) by incorporating an additional bottom-up pathway; and (**c**) BiFPN offers a superior balance between accuracy and efficiency.

BiFPN introduces bi-directional connectivity between neighboring levels of the feature pyramid, which means that information can flow from higher level features to lower level features (top-down path) or from lower level features to higher level features (bottom-up

path), which greatly enhances semantic representation and differentiation. BiFPN adopts separated convolution and lightweight design, which can improve efficiency and also reduce the number of parameters and computation. For the helmet detection task, the bidirectional feature pyramid can better solve the situation that small targets are not well recognized, which in turn leads to omission and misdetection [30].

As illustrated in Figure 5c, BiFPN is constructed upon bidirectional cross-scale connections and weighted multi-scale feature fusion. Distinct from PANet, BiFPN incorporates learnable weights to assess the significance of various input features during both top-down and bottom-up processes. The BiFPN network excludes nodes with a single input edge; such nodes are deemed insufficient for effective feature fusion due to their limited information content and reduced contribution to the feature network. Incorporating additional edges between input and output nodes within the same layer and obtaining higher-level fused features through iterative stacking enhances the feature fusion efficiency and accuracy of the BiFPN network, thereby boosting the model's capability to extract features pertinent to a specific scene. By configuring the learning parameters to ensure that the input layers are of uniform size and augmenting each feature layer with additional weights, the BiFPN network is capable of effectively merging input features of varying resolutions and evaluating their contribution to the output. The fast normalized fusion method is employed to weight the feature layer, specifically through the BiFPN weighted feature fusion technique:

$$O = \sum_i \frac{w_i \cdot I_i}{\sum_j w_j + \epsilon} \tag{7}$$

O represents the output from the BiFPN network; I_i denotes the feature from the ith layer; w_i signifies the weight associated with the ith layer feature; and w_j refers to the jth weight within the ith layer; $\epsilon = 0.001$ is used to avoid numerical instabilities, as indicated by the notation.

$$P_6^{td} = \mathrm{Conv}\left(\frac{w_1 \cdot P_6^{in} + w_2 \cdot \mathrm{Resize}\left(P_7^{in}\right)}{\varepsilon + w_1 + w_2}\right) \tag{8}$$

$$P_6^{out} = \mathrm{Conv}\left(\frac{w_1' \cdot P_6^{in} + w_2' \cdot P_6^{td} + w_3' \cdot \mathrm{Resize}\left(P_5^{out}\right)}{w_1' + w_2' + w_3' + \varepsilon}\right) \tag{9}$$

P_6^{td} denotes the feature transmitted to the sixth layer in a top-down manner; P_6^{out} represents the resultant feature of the sixth layer; Conv refers to the convolution process. P_6^{in} indicates the initial feature of the sixth layer; P_7^{in} signifies the entry feature for the seventh layer; P_5^{out} is the feature escalated within the fifth layer; Resize encompasses both upsampling and downsampling operations. The symbol ε denotes an infinitesimally small positive value; w_1 is the weighting factor for the transition from P_6^{in} to P_6^{td}; w_2 is the weighting applied to P_6^{td} following the downsampling of P_7^{in}; w_1' is the weighting for the progression from P_6^{td} to P_6^{out}; w_2' is the weighting assigned to P_6^{out}; and w_3' is the weighting for P_6^{out} subsequent to the upsampling of P_5^{out}.

3.2.3. LAHD-Head

In YOLOv8-n, a symmetric multistage compression framework is implemented, with uniform compression ratios across all detection heads. The detection head of YOLOv8 features a decoupled architecture, consisting of dual parallel streams tasked with extracting features for object classification and localization independently. Each stream is refined by a 3×3 convolutional module, adeptly capturing spatial characteristics within the image and supplying comprehensive data for subsequent categorization and positioning tasks. Following this, the output from each convolutional module is subjected to further refinement by a Conv2d module, which sharpens the feature representation and bolsters the model's object identification prowess. Ultimately, the results from these parallel tracks are utilized to calculate the classification loss (Cls loss) and the bounding box regression loss (Box loss), respectively. This decoupled approach not only optimizes the model's

detection accuracy in intricate environments but also streamlines the training process by enabling targeted optimization for distinct objectives. However, despite the decoupled design's substantial improvement in detection capabilities, it does not optimally harness the sophisticated features derived from the backbone network. The schematic of the YOLOv8-n detection head is depicted in the accompanying Figure 6.

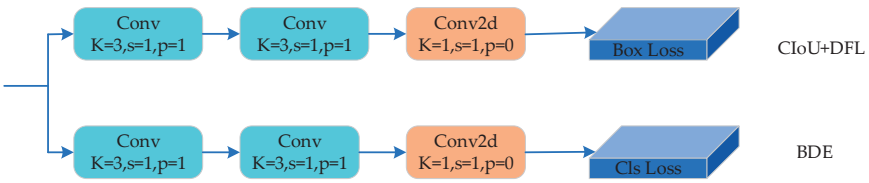


Figure 6. YOLOv8-n Detection Head.

To address the issue that conventional detection heads do not efficiently leverage the sophisticated features derived from the backbone network, this study presents an innovative, lightweight, asymmetric decoupled detection head (LADH-Head) [31]. Figure 7 illustrates the architecture of the detection head introduced in this study. A notable benefit of the LADH-Head is its implementation of 3 × 3 depthwise separable convolution (DWConv) in place of the standard 3 × 3 convolution, which significantly cuts down the parameter count and also manages to broaden the receptive field, enhance the task-specific parameters of the IOU branch, and mitigate both misdetections and false positives. The DWConv achieves a reduction in computational complexity by breaking down the standard convolution operation into depthwise convolution and pointwise convolution. The adoption of this asymmetric, lightweight detection head can substantially decrease the parameter count and enhance computational efficiency, making it ideal for environments with limited resources. Furthermore, the pointwise convolution integrates data across different channels, maintaining an adequate level of capability for feature extraction. By upgrading the detection head component, the model is now capable of higher detection accuracy while also reducing the number of model parameters and the computational resources required. This reduction in complexity and resource demands enhances computational efficiency.

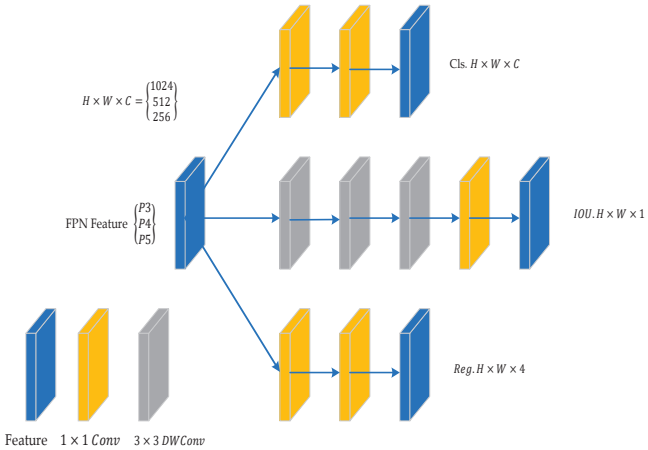


Figure 7. Network structure of the lightweight asymmetric detection head (LADH-Head).

4. Results

4.1. Experimental Environment and Dataset

The environment configuration for this experiment is shown in Table 1 below.

Table 1. Experimental environment configuration.

Name	Environmental Parameters
Operating system	Windows 11 Chinese version 64 bit
Graphics processing unit	NVIDIA GeForce RTX4050
Video memory	16
Python	3.8
Framework	PyTorch1.11

The key parameters for this network training are shown in Table 2.

Table 2. Configuration of training parameters.

Parameter	Value
Mosaic	1.0
Weight deacy	0.0005
Batch size	8
Epochs	100
Momentum	0.937
Learning rate	0.01

The publicly available dataset SHWD (Safety-Helmet-Wearing Dataset) [32–34] is used for this experiment. The SHWD dataset is a high-quality dataset focusing on helmet-wearing detection, which consists of a series of well-labeled images and is designed to train and evaluate a computer vision model on the task of helmet-detecting performance. The SHWD dataset contains 7581 images, including head objects wearing helmets (positive sample) and head objects not wearing helmets (negative sample). In this paper, the SHWD dataset is divided into a training set, a test set, and a validation set according to the ratio of 7:1:2 by random partition. After partitioning, the training set contains 5457 images, the test set contains 607 images, and the validation set contains 1517 images. The SHWD dataset contains two categories, "hat" and "person", respectively. Below is a selection of images from the SHWD dataset, as depicted in Figure 8.

Large scale Medium scale Small scale Negatives samples

Figure 8. Example of experimental dataset (Reprinted from [35]).

4.2. Evaluation of Indicators

To verify the model's performance, the main evaluation metrics include precision, mean average precision, number of parameters, FLOPs, model size, and frames per second (FPS).

$$P = \frac{TP}{TP+FP} \tag{10}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{11}$$

$$FPS = \frac{1}{Latency} \tag{12}$$

$$Parameters = (I+1) \times O = I \times O + O \tag{13}$$

where TP (true positive) denotes a positive sample correctly categorized as positive, FP (false positive) denotes a negative sample incorrectly categorized as positive, and FN (false negative) denotes a positive sample incorrectly categorized as negative. N denotes the sample category, P (precision) denotes the precision rate, and AP (average precision) denotes the average precision of a given analogy. I denotes the number of channels in the input feature map, O denotes the number of channels in the output feature map, and the +1 in the formula comes from the bias term, which has one for each output channel. The number of parameters and the amount of computation FLOPs measure the complexity of the model, while FPS measures the model's speed.

4.3. Ablation Experiments

To determine the impact of the suggested enhancements, a series of six experiments were crafted with the YOLOv8-n model serving as the foundational point for ablation analysis. The objective of these experiments was to measure the practicality and efficacy of each augmentation component, with the outcomes detailed in Table 3.

Table 3. Results of ablation experiments.

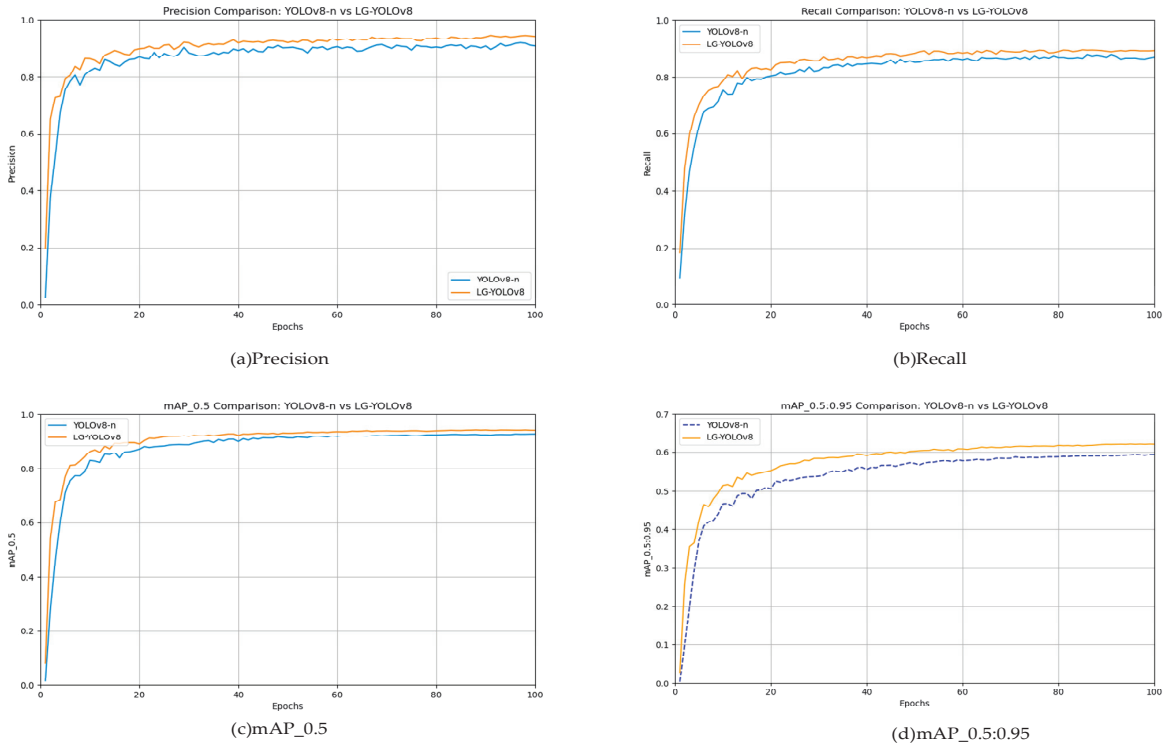
Model	G-DConv	BiFPN	LADH	Para/10^6	Model Size/M	FLOPS/G	P/%	$mAP_{0.5}$/%	FPS
Model-1	×	×	×	3.0	6.0	8.1	92.41	92.2	106.3
Model-2	√	×	×	2.2	4.6	5.7	93.4	93.7	158.6
Model-3	×	√	×	2.0	4.2	7.1	93.5	94.5	204.4
Model-4	×	×	√	2.4	4.8	5.7	91.9	94	247.1
Model-5	√	√	×	1.7	3.4	5.7	93.5	94	133.8
Model-6	√	√	√	1.2	2.7	3.7	92.2	94.1	153.3

Model-1 refers to the original YOLOv8-n, while Model-2 is the version where the C2f module is replaced with C2f-GhostDynamicConv in the YOLOv8n backbone. Compared to the baseline, Model-2 improves accuracy by 1.5%, with a 26.7% reduction in parameters and a 29.2% drop in computation. This shows that the replacement model has a more significant impact on parameter reduction. Model-3 introduces a bidirectional feature pyramid in the feature fusion network, slightly increasing computation but boosting accuracy by 2.3%, reducing parameters by 33.3%, and decreasing model size by 30% compared to Model-2. This demonstrates that the bidirectional feature pyramid excels at feature extraction and model optimization through separable convolution and a lightweight design. Model-4 replaces the baseline detection head with the lighter LADH-Head, which does not improve accuracy or model size over Model-3 but enhances frames per second by 20.9%. Model-5 adds a bidirectional feature pyramid to Model-2, resulting in a slight accuracy increase, a 22.7% reduction in parameters, and a 26% decrease in model size. Model-6 builds on Model-5 by adding LADH-Head, showing a 1.9% accuracy improvement compared to Model-1, with the model size reduced to 55% of the original and computation cut by 54.3%. These experiments show that the optimized YOLOv8n algorithm not only maintains high accuracy in helmet detection but also significantly reduces parameters, computation, and model size.

4.4. Model Training Analysis Before and After Improvement

Figure 9 shows the changes in key metrics of the YOLOv8-n and LG-YOLOv8 models throughout the training process, further demonstrating the superiority of the improved models. Figure 9a–d depicts the fluctuations of precision, recall, mAP_0.5, and mAP_0.5:0.95, respectively, for the training cycle (epoch = 100). The comparison of the accuracies clearly shows that the accuracy of LG-YOLOv8 is always higher than that of YOLOv8-n throughout the training process, which indicates that the model has been improved. By observing the recall rate, the recall rate increases rapidly in the first 40 epochs. And in the whole training process, the recall rate of LG-YOLOv8 is always higher than

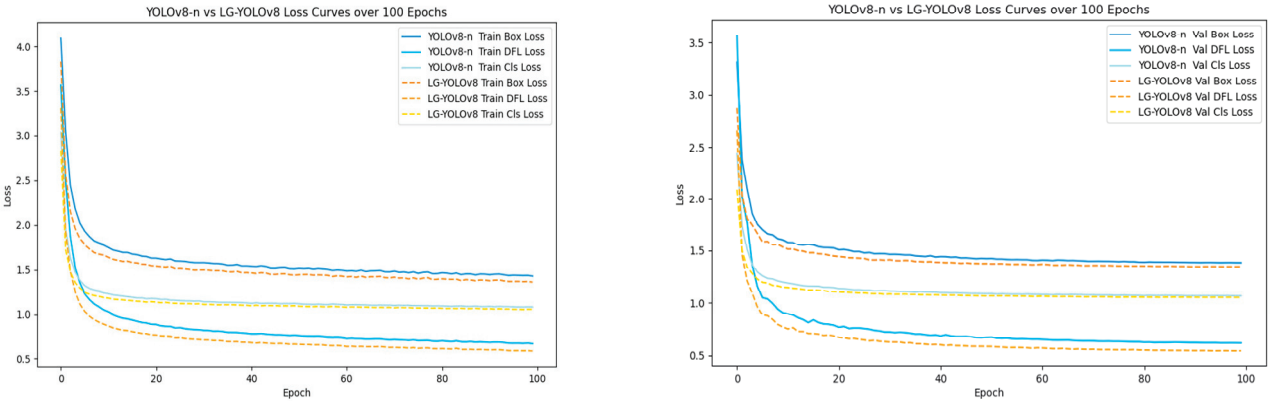
that of YOLOv8-n, which indicates that the improved model has a lower miss rate and can better detect all positive samples. The mAP_0.5 curve shows that LG-YOLOv8 is higher than YOLOv8-n in the whole training process, which indicates that the improved model has stronger robustness for safety helmet detection and can detect and identify targets more effectively. It can be observed from mAP_0.5:0.95 that the enhanced model demonstrates more superior overall detection performance under diverse matching circumstances throughout the entire 100 rounds of training. Figure 10 presents the variations of different types of losses of the YOLOv8-n and LG-YOLOv8 models during training and validation. The loss functions in Figure 10a,b indicate that the enhanced model converges more rapidly and has better performance in inference. In summary, the superiority of LG-YOLOv8 in all indicators indicates that optimized modules can effectively improve the overall performance of the model, further highlighting the higher accuracy and reliability of LG-YOLOv8 in safety helmet detection.



(a)Precision (b)Recall

(c)mAP_0.5 (d)mAP_0.5:0.95

Figure 9. Changes in key metrics during YOLOv 8-n and LG-YOLOv8 trainings.



(a)Training set loss function (b)Validation set loss function

Figure 10. Changes in loss during YOLOv8-n and LG-YOLOv8 training.

4.5. Comparative Experiments

To underscore the effectiveness and practicality of the LG-YOLOv8 algorithm for compact studies, we compared its performance to leading algorithms in helmet detection. The comparative analysis and respective data are elaborated in Table 4.

Table 4. Comparison results of different models.

Model	Parameters/10^6	FLOPS/G	P/%	$mAP_{0.5}$/%
SSD [3]	12.3	28.7	-	90.2
YOLOv3-Tiny	8.7	12.9	-	61.2
YOLOv3	61.5	154.9	-	89.1
YOLOv5s	7.1	15.8	93.6	90.9
YOLOv7	37.2	105.1	93.3	93.4
ML-YOLOv3 [36]	18.1	46.0	-	86.5
MG-YOLOv5s [37]	4.06	67	-	95.4
YOLOv8-SLIM-CA [38]	2.74	11.3	93.8	94
YOLOv8n-ASF-DH [39]	5.01	11.8	94	95.6
LG-YOLOv8	1.2	3.7	92.2	93.4

Table 4 reveals that LG-YOLOv8 outperforms SSD in average accuracy by a margin of 3% while substantially reducing parameter count and computational load. When juxtaposed with YOLOv5 and YOLOv7, our model demonstrates a commendable lightweight effect. In comparison to enhanced variations of YOLOv3 and YOLOv3-Tiny, the refined algorithm of this study shows marked enhancements in average accuracy, with a pronounced decrease in computational demand and parameter requirements. Further, against YOLOv5-MN, LG-YOLOv8 exhibits a marginal uptick in average accuracy with a notable 59.3% reduction in computation. Against three advanced YOLOv8 iterations—YOLOv8-Slim-CA, FEV-YOLOv8n, and YOLOv8-S-LE—the algorithm developed herein stands out, especially in parameter number and computation efficiency. The current model's strength is not just in maintaining high precision but also in optimizing resource usage, making it highly suitable for edge device deployment. The effectiveness of LG-YOLOv8 is visually corroborated by Figure 11, highlighting the algorithm's advantage with greater clarity.

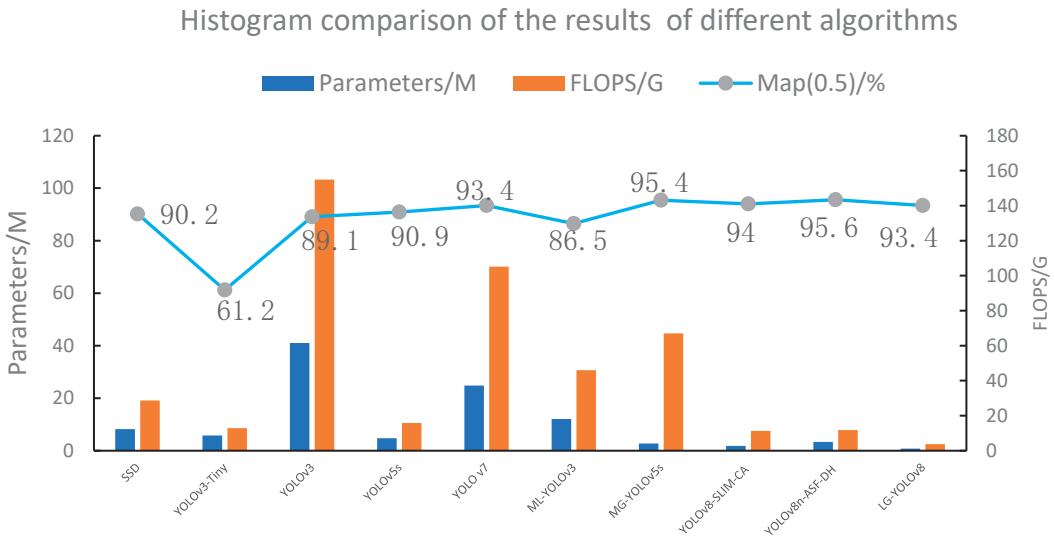


Figure 11. Histogram comparison of results of different algorithms.

4.6. Visualization of the Results

To offer a clearer evaluation of the enhanced YOLOv8-n model's detection capabilities, we chose safety helmet detection in diverse challenging environments from the SWHD dataset. The selected original images pose challenges in identifying small objects, handling

crowded scenes, and managing complex scenarios. As shown in Figure 12, our proposed method not only enhances the feature extraction ability but also effectively addresses missing cases in some complex scenes and small target cases.

(a) original image (b) YOLOv8-n (c)Ghost-LADH YOLOv8

Figure 12. Visualization results for different scenarios (adapted from ref. [35]). (**a**) The pictures in the original dataset and the helmet detection pictures in different scenarios; (**b**) Base model YOLOv8-n; (**c**) Improved model LG-YOLOv8.

5. Conclusions

To counteract the deficiencies of the standard YOLOv8-n algorithm and address external factors that can affect helmet detection performance, this manuscript presents an advanced model known as LG-YOLOv8. Constructed upon the YOLOv8-n foundation, this model integrates a triad of architectural enhancements. These improvements notably decrease the computational load without sacrificing accuracy, aligning the model more closely with real-world usage scenarios. The introduction of the C2f-GhostDynamicConv module serves to refine the original C2f module, leading to more efficient use of computational resources and better performance in low-precision floating-point operations. Moreover, the paper presents a revamped feature extraction segment and detection head, complete with a bidirectional feature pyramid that amplifies the semantic data extracted from helmet images, consequently bolstering detection capabilities. The novel lightweight asymmetric detection head, termed LADH-Head, minimizes model parameters and accelerates detection without compromising the integrity of the classification and regression processes initially challenged by the traditional coupled head design. Empirical evidence supports that LG-YOLOv8 is adept at fulfilling helmet detection requirements with robust practicality, achieving a commendable equilibrium between detection proficiency and computational expenditure. It outperforms its contemporaries in specific performance metrics and is better suited for deployment on resource-constrained edge devices. While the algorithm has demonstrated impressive outcomes, there remains ample opportunity for refinement. In subsequent research, we aim to further unearth the untapped potential of the YOLO model for helmet detection. Additionally, we plan to delve into more potent

algorithms for the detection and tracking of smaller targets to augment their recognition precision. We intend to investigate the adaptability of our method to other object detection algorithms to harvest superior performance. Moreover, we will seek to amass a more extensive and diverse dataset to fortify the models' generalization capabilities, rendering them more potent and adaptable in intricate industrial settings.

Author Contributions: Conceptualization, Y.W. and Z.F.; methodology, Z.F.; formal analysis, Y.W. and Z.F.; investigation, Y.W., M.C. and W.L.; resources, Y.W. and Z.F.; data curation, Y.W.; writing—original draft preparation, Z.F.; writing—review and editing, Y.W. and Z.F.; visualization, Z.F.; supervision, Y.W. and Z.F.; project administration, Z.F. and Z.Q.; funding acquisition, Z.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Heilongjiang Postdoctoral Fund to pursue scientific research (grant number LBH-Z23025), Heilongjiang Province Colleges and Universities Basic Scientific Research Business Expenses Project (grant number 2023-KYYWF-1052), Harbin University of Commerce Industrialization Project (grant number 22CZ04), Collaborative Innovation Achievement Program of Double First-class Disciplines in Heilongjiang Province (grant number LJGXCG2022-085).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article. The SHWD dataset can be download in https://github.com/njvisionpower/Safety-Helmet-Wearing-Dataset (accessed on 22 July 2024). Additional original contributions presented in the study are included in the article, and further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
2. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
3. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings, Part I 14. Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37.
4. Xu, X.; Jiang, Y.; Chen, W.; Huang, Y.; Zhang, Y.; Sun, X. DAMO-YOLO: A Report on Real-Time Object Detection Design. *arXiv* **2022**, arXiv:2211.15444.
5. Adarsh, P.; Rathi, P.; Kumar, M. YOLO v3-Tiny: Object Detection and Recognition using one stage improved model. In Proceedings of the 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 6–7 March 2020; pp. 687–694.
6. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W.; et al. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. *arXiv* **2022**, arXiv:2209.02976.
7. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 7464–7475.
8. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. YOLOX: Exceeding YOLO Series in 2021. *arXiv* **2021**, arXiv:2107.08430.
9. Chen, Q.; Wang, Y.; Yang, T.; Zhang, X.; Cheng, J.; Sun, J. You Only Look One-level Feature. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Virtual, 19–25 June 2021; p. 13034.
10. Glenn, J.; Ayush, C.; Jing, Q. Ultralytics YOLOv8. Available online: https://github.com/ultralytics/ultralytics (accessed on 2 July 2024).
11. Lian, Y.; Li, J.; Dong, S.; Li, X. HR-YOLO: A Multi-Branch Network Model for Helmet Detection Combined with High-Resolution Network and YOLOv5. *Electronics* **2024**, *13*, 2271. [CrossRef]
12. Liang, H.; Seo, S. Automatic detection of construction workers' helmet wear based on lightweight deep learning. *Appl. Sci.* **2022**, *12*, 10369. [CrossRef]
13. Chen, S.; Tang, W.; Ji, T.; Zhu, H.; Ouyang, Y.; Wang, W. Detection of safety helmet wearing based on improved faster R-CNN. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–7.
14. Chen, J.; Deng, S.; Wang, P.; Huang, X.; Liu, Y. Lightweight helmet detection algorithm using an improved YOLOv4. *Sensors* **2023**, *23*, 1256. [CrossRef] [PubMed]

15. Ma, N.; Zhang, X.; Zheng, H.T.; Sun, J. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In Proceedings of the European Conference on Computer Vision–ECCV 2018, Munich, Germany, 8–14 September 2018; pp. 116–131.
16. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
17. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. Ghostnet: More features from cheap operations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 1580–1589.
18. Talaat, F.M.; ZainEldin, H. An improved fire detection approach based on YOLO-v8 for smart cities. *Neural Comput. Appl.* **2023**, *35*, 20939–20954. [CrossRef]
19. Li, Y.; Fan, Q.; Huang, H.; Han, Z.; Gu, Q. A Modified YOLOv8 Detection Network for UAV Aerial Image Recognition. *Drones* **2023**, *7*, 304. [CrossRef]
20. Wang, C.Y.; Liao, H.Y.M.; Wu, Y.H. Cspnet: A new backbone that can enhance learning capability of cnn. In Proceedings of the IEEE 2020/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 390–391.
21. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
22. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
23. Ma, N.; Su, Y.; Yang, L.; Li, Z.; Yan, H. Wheat Seed Detection and Counting Method Based on Improved YOLOv8 Model. *Sensors* **2024**, *24*, 1654. [CrossRef] [PubMed]
24. Yang, B.; Bender, G.; Le, Q.V.; Ngiam, J. Condconv: Conditionally parameterized convolutions for efficient inference. *Adv. Neural Inf. Process. Syst.* **2019**, *32*. Available online: https://neurips.cc/virtual/2019/poster/13351 (accessed on 26 October 2024).
25. Han, K.; Wang, Y.; Guo, J.; Wu, E. ParameterNet: Parameters Are All You Need for Large-scale Visual Pretraining of Mobile Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 15751–15761.
26. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
27. Kim, J.H.; Kim, N.; Park, Y.W.; Won, C.S. Object detection and classification based on YOLO-V5 with improved maritime dataset. *J. Mar. Sci. Eng.* **2022**, *10*, 377. [CrossRef]
28. Qu, Z.; Gao, L.Y.; Wang, S.Y.; Yin, H.N.; Yi, T.M. An improved YOLOv5 method for large objects detection with multi-scale feature cross-layer fusion network. *Image Vis. Comput.* **2022**, *125*, 104518. [CrossRef]
29. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 10781–10790.
30. Guo, C.; Fan, B.; Zhang, Q.; Xiang, S.; Pan, C. Augfpn: Improving multi-scale feature learning for object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 12595–12604.
31. Zhang, J.; Chen, Z.; Yan, G.; Wang, Y.; Hu, B. Faster and Lightweight: An Improved YOLOv5 Object Detector for Remote Sensing Images. *Remote Sens.* **2023**, *15*, 4974. [CrossRef]
32. Wang, J.; Zhu, G.; Wu, S.; Luo, C. Worker's helmet recognition and identity recognition based on deep learning. *Open J. Model. Simul.* **2021**, *9*, 135–145. [CrossRef]
33. Yue, S.; Zhang, Q.; Shao, D.; Fan, Y.; Bai, J. Safety helmet wearing status detection based on improved boosted random ferns. *Multimed. Tools Appl.* **2022**, *81*, 16783–16796. [CrossRef]
34. Sun, H.; Gong, P. November. A Safety-Helmet Detection Algorithm Based on Attention Mechanism. In Proceedings of the 2021 7th IEEE International Conference on Network Intelligence and Digital Content, Beijing, China, 17–19 November 2021; pp. 11–15.
35. njvisionpower. Safety-Helmet-Wearing-Dataset. 2019. Available online: https://github.com/njvisionpower/Safety-Helmet-Wearing-Dataset (accessed on 22 July 2024).
36. Deng, L.; Li, H.; Liu, H.; Gu, J. A lightweight YOLOv3 algorithm used for safety helmet detection. *Sci. Rep.* **2022**, *12*, 10981. [CrossRef] [PubMed]
37. Xiao, Z.; Liu, W.; Ye, Z.; Yuan, J.; Liu, S. MG-YOLOv5s: A Faster and Stronger Helmet Detection Algorithm. *Comput. Syst. Sci. Eng.* **2024**, *48*, 1009–1029. [CrossRef]
38. Lin, B. Safety helmet detection based on improved YOLOv8. *IEEE Access* **2024**, *12*, 28260–28272. [CrossRef]
39. Lin, B. YOLOv8n-ASF-DH: An Enhanced Safety Helmet Detection Method. *IEEE Access* **2024**, *12*, 126313–126328. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

urban science

MDPI

Article

Index for Assessing the Performance Level of Vehicular Traffic on Urban Streets

Ernesto De la Cruz-Nicolás, Hugo Estrada-Esquivel, Alicia Martínez-Rebollar *, Odette Alejandra Pliego-Martínez and Eddie Clemente

CENIDET, National Technological of Mexico, Cuernavaca 62490, Mexico;
d21ce090@cenidet.tecnm.mx (E.D.l.C.-N.); hugo.ee@cenidet.tecnm.mx (H.E.-E.);
d21ce092@cenidet.tecnm.mx (O.A.P.-M.); eddie.ct@cenidet.tecnm.mx (E.C.)
* Correspondence: alicia.mr@cenidet.tecnm.mx

Abstract: Vehicular transportation is one of the most widely used modes in modern cities for reaching key destinations such as workplaces, healthcare facilities, recreational areas, and educational institutions, among others. However, the performance of vehicular traffic on these roads can vary significantly due to the influence of different environmental variables. In the literature, factors such as traffic incidents, weather conditions, road infrastructure, and driving habits, among others, have been identified as impacting vehicular traffic performance. In this context, predictive models have been developed to anticipate congestion at specific points in cities based on statistics, machine learning, simulation, and complex networks. This study proposes a novel index aimed at assessing the level of performance of vehicular traffic on streets based on the relationship among relevant urban environmental variables. This index is generated through Genetic Programming, considering a set of variables related to traffic, incidents, and services. The case study will focus on the streets of the Tlalpan Municipality in Mexico City.

Keywords: index; assessment; vehicular traffic; performance level; variables

1. Introduction

Vehicular transportation in current large cities contributes to reach services such as healthcare facilities, workplaces, culture services, and education institutions, among others, as described by the Center for Complexity Sciences [1]. City traffic includes multiple forms of mobility, such as walking, cycling, public transport, shared vehicles, private vehicles, motorcycles, and other means, according to the urban mobility pyramid [2].

Among these modes, private transportation has become the most commonly used for point-to-point travel. This is evidenced by the growing number of registered vehicles reported by state and municipal governments in Mexico—a trend historically documented by the National Institute of Statistics and Geography [3].

The increase in the number of vehicles in cities, combined with factors such as population density, poor urban planning, inadequate road infrastructure, and the unequal distribution of essential services, contributes to fluctuating levels of vehicular traffic on certain streets. As noted in the studies by Mondal [4] and Sun [5], the occurrence of traffic incidents in specific regions also exacerbates the situation. Furthermore, temporal and spatial factors, such as the increase in traffic accidents, influence the performance of vehicular traffic, as highlighted by Wang's research [6].

Traffic policies, driving culture, and economic factors, according to Almeida [7], also influence vehicular traffic performance. Additionally, urban infrastructure plays a crucial role in shaping traffic patterns, as mentioned by Nian [8]. Various solutions have been proposed to enhance traffic performance in cities, with traffic modeling being one of the earliest approaches. According to Kushchenko [9], this type of modeling enables analysis

of vehicle movement and identifying key variables, such as intersections, queue dynamics and lengths, speed, and other characteristics.

Sun's analysis [10] highlights common variables in traffic models, such as speed, vehicle length, distances, number of vehicles per kilometer, flow rate, intersections, traffic signals, and road segment lengths. However, these models only allow the analysts to visualize vehicular flow under normal conditions. In situations when traffic density increases, the phenomenon of congestion arises, a situation that can also be attributed to other characteristics of the urban environment. Studies have addressed congestion through indicators or indices that characterize it, generally relating average speed to traffic density.

This research makes a significant contribution through the development of a novel mathematical model that introduces an index to evaluate vehicular traffic performance on urban streets. While Genetic Programming and symbolic regression are not entirely new techniques, their application in this specific context brings innovative value by enabling the discovery of complex relationships among multiple variables that impact traffic. These techniques are employed because they offer flexibility and the capacity to generate models that adapt to the nonlinear and dynamic nature of urban traffic, something that is not always achievable with traditional or exact approaches.

The need for a new indexing method arises from the limitations of existing approaches, which often fail to capture the complexity and variability of urban traffic environments. Current indices tend to oversimplify reality by focusing only on speed indicators, overlooking critical factors that truly affect traffic, such as incidents, infrastructure, and surrounding services. Although sets of variables that influence traffic flow have been identified in the literature, models associating these variables to adequately characterize vehicular traffic performance have not been developed. This lack of a comprehensive approach may lead to an inadequate assessment of traffic performance, especially in areas where multiple influences interact significantly.

The choice of Genetic Programming and symbolic regression addresses the need for an approach that does not rely on rigid assumptions about relationships between variables. Unlike exact methods such as numerical approaches or traditional regression techniques, which require predefined models and, in many cases, linearity, Genetic Programming enables the evolutionary generation of solutions by exploring a broader space of potential mathematical formulas that fit the data. This approach is particularly useful when variable interactions are not evident or when there is a need to simultaneously optimize multiple objective functions.

On the other hand, although artificial intelligence (AI) techniques such as neural networks or machine learning models can be effective in certain scenarios, they often act as "black boxes," making it difficult to interpret the results. In this case, symbolic regression enables the generation of explicit models that not only predict traffic behavior but are also interpretable, facilitating the understanding of how variables related to traffic, incidents, and services interact.

The proposed index incorporates traffic, incident, and service (points of interest) variables, allowing for an index that more accurately characterizes the performance or equilibrium state of vehicular traffic on a street. This combination of factors is key to better capturing the changing and complex conditions of urban streets, providing a model that is not only predictive but also more attuned to the realities of urban mobility.

Thus, the proposed index offers an integral evaluation of traffic performance based on the interaction between the independent and dependent variables involved, providing a more accurate, interpretable, and adaptive solution to the fluctuating conditions of urban traffic.

2. Related Works

The works reviewed in the literature are grouped into two main approaches. (i) research that develops mathematical models to evaluate traffic congestion based on various vari-

ables and (ii) research that identifies the variables with the greatest impact on vehicular congestion and also analyzes how they affect mobility in urban environments.

Equations, functions, and mathematical models have been proposed in several studies to characterize traffic congestion through the relationship between different variables. For example, Mohan, He and Liponhay [11–13] proposes a Zone Speed Index (IVZ) aimed at classifying areas as fast or slow based on the relationship between those zones.

Pandey [14] addresses the heterogeneity of mixed traffic in road networks, proposing a Heterogeneity Index (HI) that quantifies the diversity of vehicle categories in the traffic flow. A high HI value reflects greater vehicular diversity, which favors mobility, based on Passenger Car Units (PCU) for different types of vehicles and their proportion within the flow.

On the other hand, Leitner [15] proposes a speed performance index to assess congestion in the road network, developing congestion indices for both individual road segments and the entire network. These indices are based on variables such as average travel speed, the proportion of non-congested time, and segment length.

Another relevant approach is the work of Seong [16], who introduces new metrics through Hägerstrand's space–time cube to characterize traffic congestion, using distance and time as the key variables.

Wang [17] proposes three methods to calculate the congestion index: the first based on travel speed, the second on roadway saturation degree, and the third on integral parameters such as travel efficiency and low-speed proportion.

On the other hand, Arirja [18] develops an index to represent congestion density over time, introducing the occupancy variable, which combines speed and vehicle length to measure the time in which a road segment is occupied.

Liu [19] constructs a prediction model for congestion time through a multivariable equation that includes traffic density, average delay time, traffic flow, and peak hours as independent variables.

The analysis of the literature also identified variables influencing traffic congestion. For example, Chen [20] highlights traffic accidents and response time as key factors. Bian [21] identifies variables such as per capita street area and vehicle ownership, while Mahona [22] emphasizes the importance of continuous crossings, bus stops, and traffic lights. Yu [23], for his part, underscores intersections as one of the main causes of congestion.

Points of interest in cities, such as shopping centers and schools, also play an important role, according to Gullotta [24], as they attract vehicles during specific time periods, increasing congestion. Rahman [25] identifies income and employment clustering in specific areas as a factor exacerbating traffic, while Pi [26] mentions road design, incidents, and poor driving habits as contributing factors to congestion.

It is fascinating to explore the different indices and indicators, like the indicators in [27] that propose important metrics for evaluating alternative transportation systems—the Mode Efficiency Factor and the Passenger Transportation Supply Efficiency Index—that allow for the evaluation of traffic congestion. This research work presents an innovative proposal: an index that integrates variables from accessible sources, offering a more accurate visualization of mobility performance. This index aims to provide a more comprehensive understanding of the problem of congestion in urban environments and to provide a solid foundation for addressing it.

3. Methodology for the Creation of an Index to Assess the Performance Level of Vehicular Traffic on Urban Streets

The methodology proposed to create an index to evaluate vehicular traffic performance on urban streets is divided into three main phases. The first phase addresses the case study, providing a detailed description of the geographical area from which data on traffic, incidents, and services are collected. The second phase focuses on data acquisition, outlining the tools used to gather this information. Finally, the third phase describes the methodological approach used in designing the proposed mobility index.

3.1. Case Study Definition Phase

The case study defined for this research is the vehicular traffic in the area of Tlalpan Municipality, located in Mexico City, which covers the largest territorial expanse in the south-southwest part of the metropolis. Tlalpan is bordered to the north by the Municipalities of Álvaro Obregón and Coyoacán, to the east by Xochimilco and Milpa Alta, to the south by Huitzilac in Morelos, and to the west by Magdalena Contreras, as well as by the municipalities of Tianguistenco and Jalatlaco of the State of Mexico. With an area of 306.52 km^2, Tlalpan represents 20% of the total area of Mexico City, ranking first compared to the other Mexico City Municipalities, as reported by the National Institute of Statistics and Geography [28]. This municipality is characterized by having both urban and rural areas, providing a diverse environment that facilitates making inferences and formulating conclusions regarding the findings of our research.

3.2. Data Acquisition Phase

The data collection consisted of obtaining the variables shown in Table 1, grouped into three categories: traffic, incidents, and services (points of interest), each of which plays an essential role in traffic performance on the streets. Traffic variables, such as jamFactor, describe the level of congestion, while Speed and Speed_Uncapped allow for a comparison between expected speed under optimal conditions and actual speed, providing a direct indicator of vehicle flow efficiency. Variables like Length and Number_Segments detail the capacity of the roadway, and temporal factors, such as Work_Day_Traffic, Traffic_Block_Type, and Traffic_Hour, facilitate the identification of congestion patterns associated with the time of day, day of the week, or whether it is a workday, enabling the assessment of traffic performance in different temporal contexts.

Table 1. Detailed description of the number of records and variables in the traffic-incident-POIs dataset.

Dataset	Rows	Number of Variables
Tráfico-incidencias-POIs	873,020,000	35

On the other hand, incident variables such as Type and Criticality reflect the type and severity of events, such as accidents or road closures, that directly affect circulation and decrease traffic speed in specific areas. Variables like Start_Time, End_Time, and Incident_Block_Type provide information on the duration and frequency of incidents, helping to predict their patterns and understand their impact on road accessibility. Finally, the points of interest (POIs) variables, represented by Name and Bbox_POI, identify the location of services and destinations that generate higher traffic in their surroundings. The concentration of POIs in an area increases the likelihood of congestion, as these sites attract a significant flow of vehicles at specific times.

To deepen the analysis of traffic congestion, it would be desirable to incorporate a broader range of critical factors, such as weather, road conditions, and driver behavior, which are highly likely to influence vehicle density and traffic flow. These elements can introduce significant variations in traffic, providing a richer and more accurate understanding of road dynamics in various scenarios. However, these factors were not included in this study due to limitations related to the availability and consistency of the necessary data. Historical and real-time information about weather or specific road conditions is not always available for each analyzed segment, and accurately obtaining these conditions would require exhaustive and continuous data collection, which is often impractical in many urban contexts. Furthermore, the variability of these factors over time and space makes their integration into the analysis complex and potentially biased, as homogeneous and replicable measurements cannot be guaranteed.

In line with these challenges, the methodology focused on high-impact variables whose collection and measurement are consistent over time, thus allowing for a robust

and replicable analysis of vehicular congestion. This approach excludes personal and contextual factors that are difficult to measure, such as driver stress or the detailed weather conditions of each area, which, while critical, are challenging to obtain in representative and homogeneous volumes without affecting the viability of the study. By prioritizing variables whose influence is directly measurable on traffic performance, the study evaluates congestion without resorting to invasive or complex data acquisition methods. Thus, the relevance and feasibility of the analysis are optimized, maintaining the methodological rigor necessary to effectively assess the impact of traffic on road performance despite not encompassing all possible external factors.

Data acquisition on traffic and road incidents was carried out using the Here Maps Application Programming Interface (API), following the methodology proposed by De la Cruz-Nicolás [29]. There are other alternatives for acquiring traffic data, such as the one used in the work of [30]. Comprehensive traffic and incident data were collected from October 2023 to March 2024, covering 24 h a day with requests made every 5 min. The Geoapify API was used to obtain data related to points of interest. The datasets obtained on traffic, road incidents, and services are presented in Table 1.

The description of each variable in the Traffic-incident-POIs dataset is shown in Table 2.

Table 2. Description of the variables in the Traffic-incident-POIs dataset.

Dataset	ID	Variable Name	Variable Description
-	-	jamFactor	Describe traffic congestion.
Traffic	X1	Bbox_Traffic	Bounding box where traffic originates.
	X2	Confidence	Describe whether the data is real-time or historical.
	X3	Free_Flow	Speed without traffic.
	X4	Length	Length of the segment.
	X5	Number_Segments	Number of segments that the monitored street contains.
	X6	Speed	Expected speed.
	X7	Speed_Uncapped	Current speed.
	X8	Daily_Traffic	Day of traffic monitoring.
	X9	Day_Off_Traffic	Value of 1 when the day is a holiday and 0 when it is not.
	X10	Traffic_Hour	Traffic monitoring time, the format is HH: MM.
	X11	Traffic_Month	Month of traffic monitoring.
	X12	Traffic_Day_Number	Day number of traffic monitoring (1 = Monday, 2 = Tuesday, . . . 7 = Sunday).
	X13	Traffic_Minute	Minute of traffic monitoring.
	X14	Work_Day_Traffic	Traffic monitoring: workweek = 1, weekend = 0.
	X15	Traffic_Block_Type	Traffic by time blocks (block 1 = 0–3 h, block 2 = 4–7 h, block 3 = 8–11 h, . . . block 6 = 20–23 h).
Incident	X16	Bbox_Incident	Bounding box where the incident originates.
	X17	Road_Closed	Indicates if the street is closed due to the incident (0 open and 1 closed).
	X18	Criticality	Severity of the incident.
	X19	Type	Type of incident: accident, construction, congestion, weather, among others.
	X20	Description	Provide a description of the location of the incident.
	X21	Olr	ID of the shape of the incident.
	X22	Start_Time	Start time of the incident.
	X23	End_Time	End time of the incident.
	X24	Incident_Day	Day of the incident monitoring (values between 1 and 30, or 1 and 31 depending on the month).
	X25	Day_Off_Incident	Value of 1 when the day is a holiday and 0 when it is not for incident monitoring.
	X26	Incident_Hour	Hour of incident monitoring.
	X27	Incident_Month	Month of incident monitoring.
	X28	Incident_Minute	Minute of incident monitoring.
	X29	Incident_Day_Number	Day number of incident monitoring (1 = Monday, 2 = Tuesday, . . . 7 = Sunday).
	X30	Incident_Block_Type	Time block in which the incident was recorded.
	X31	Work_Day_Incident	Incident monitoring during the workweek = 1 and on weekends = 0.
	X32	Category_Number	Incident category.
	X33	Time	Time in minutes, for example, 0:00 h minute 0, 0:05 h minute 5, 0:10 h minute 10, and so on up to 23:55 h minute 1435.
Points of interest	X34	Name	Description of the point of interest.
	X35	Bbox_POI	Bounding box of the point of interest.

3.3. Index to Evaluate the Performance of Vehicular Traffic Developing Phase

The proposed approach for developing an index to evaluate the performance level of vehicular traffic on urban streets has been divided into five stages. The process begins with the collection of data on traffic, traffic incidents, and points of interest, and they conclude with the generation of the index as a result, as illustrated in Figure 1.

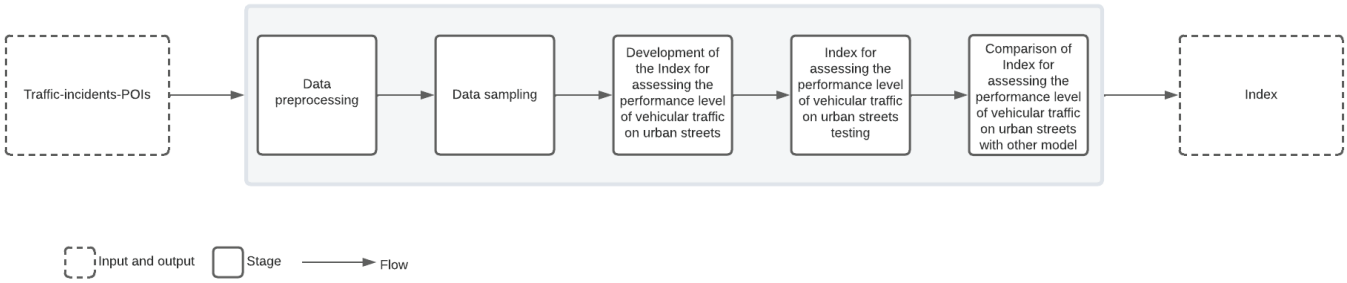


Figure 1. Stages that allow for obtaining the index to evaluate the performance level of vehicular traffic on urban streets.

In the first stage, the preprocessing of traffic data, traffic incidents, and points of interest is carried out, which includes transformation, standardization, and selection of the most relevant variables. In the second stage, a sample is extracted from the preprocessed dataset of Traffic-Incidents-POIs to expedite the analysis and modeling of the index.

The third stage involves the development of a mathematical model that will represent the index. This is done by using genetic programming and the Traffic-Incidents-POIs dataset as input. The fourth stage consists of conducting tests of the index that evaluates traffic congestion, employing a new dataset and statistical metrics to measure the accuracy of the model in assessing the level of performance of vehicular traffic.

Finally, the fifth stage focuses on comparing the obtained index with other indices that can be found in the literature for assessing the performance of vehicular traffic. The following sections provide a detailed explanation of the proposed stages.

3.3.1. Stage 1: Data Preprocessing

Data preprocessing is established as the fundamental basis for ensuring the effectiveness and accuracy of the final results. In this research, the data preprocessing process was carried out through the following three steps: data transformation, standardization, and relevant variable selection

Step 1. Data Transformation: The primary objective of data transformation is to adjust the distributions of the data to resemble a normal form, which is essential in model construction. To assess whether the variables influencing traffic performance exhibit normal behavior, the Kolmogorov-Smirnov test was applied. Performing this test on the Traffic-Incidents-POIs Sample dataset provided valuable information about the distribution of variables related to traffic performance, revealing that these did not follow a normal distribution. This finding indicates the need to implement appropriate transformations to address this irregularity in the data and ensure the validity of subsequent analyses.

To determine the type of transformation to apply to the Traffic-Incidents-POIs dataset, an analysis was conducted on incidents across various streets. The data exhibited a right-skewed distribution, characterized by a higher number of streets with few incidents, while some streets had a significantly higher count. Additionally, vehicle speeds on most streets showed an asymmetric distribution, with low speeds prevailing but isolated cases of vehicles traveling at very high speeds, also resulting in a right-skewed distribution. On the other hand, points of interest (POIs) located on these streets are represented in different scales and units, adding an additional layer of complexity to the data analysis. In this context, the following transformations are justified: the square root, which is useful for count data with a right skew, such as traffic incidents, as it reduces variance and

brings the distribution closer to normality; the cubic transformation, ideal for data with moderate skewness, such as vehicle speeds, which helps smooth the tails and improve distribution symmetry; and Z-score normalization, which is especially applicable when comparing variables on different scales, such as the performance scores of POIs, allowing data standardization and facilitating direct comparison between disparate variables. The applied transformation included the implementation of square root transformation, z-score normalization, and cubic transformation, using Equations (1)–(3) on the Traffic-Incidents-POIs dataset to normalize the data.

$$\text{Normalized variable} = \sqrt{Original\ variable} \tag{1}$$

$$z = \frac{x - \mu}{\sigma} \tag{2}$$

where:

x It is the individual value of the variable,

μ It is the mean of the variable's distribution,

σ It is the standard deviation of the variable's distribution.

$$\text{Normalized variable} = \sqrt[3]{Original\ variable} \tag{3}$$

The result of normalizing the Traffic-Incidents-POIs dataset has demonstrated that applying the cubic transformation significantly improves the normality of this dataset.

Step 2. Standardization: For the Traffic-Incidents-POIs dataset, which has variables on different scales, a standardization was performed that adjusts the values within a range of 0 to 1. This process was carried out using Equation (4).

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{4}$$

where

x It is the original value that is to be normalized,

x_{min} It is the minimum value in the original dataset,

x_{max} It is the maximum value in the original dataset.

Standardization allowed all variables in the Traffic-Incidents-POIs dataset to be adjusted to a range between 0 and 1.

Step 3. Variable Selection: To identify the most significant variables in the Traffic-Incidents-POIs dataset, the rotation matrix from the Principal Component Analysis (PCA) was used. This approach allowed for the determination of variables that have a notable impact on vehicle traffic performance. The most relevant variables are presented in Table 3.

Table 3. Most Relevant Variables from the Traffic-Incidents-POIs Dataset Tráfico-Incidencias-POIs.

Number	Variable	Number	Variable	Number	Variable
1	Length	9	Olr	17	Speed
2	Start_Time	10	Category_Number	18	Incident_Minute
3	Incident_Month	11	Description	19	Traffic_Minute
4	Criticality	12	Traffic_Block_Type	20	jamFactor
5	Number_Segments	13	Type	21	Bbox_POI
6	Road_Closed	14	Bbox_Traffic	22	Daily_Traffic
7	Incident_Hour	15	End_Time	23	Incident_Day
8	Traffic_Hour	16	incidencia_tipobloque		

3.3.2. Stage 2: Data Sampling

The dataset used in this research, named Traffic-Incidents-POIs, contains a total of 873,020,000 records. To obtain a representative sample of 1,000,000 records, we implemented simple random sampling, following these steps:

Step 1. Population Definition: The dataset of 873,020,000 records was defined as the population.

Step 2. Simple Random Sampling: To select the sample, we used the simple random sampling function from Python's Random library. This method ensures that each record in the population has an equal probability of being selected, expressed in Equation (5):

$$P(i) = \frac{1}{N} \tag{5}$$

where $P(i)$ is the probability of selecting record i, and N is the total population size (873,020,000 in this case).

Step 3. Selection Process: Selection was performed by generating a random number that identifies the positions of records within the original dataset. Through this process, we extracted 1,000,000 records randomly, without replacement, meaning that once a record was selected, it could not be selected again. This approach is essential to avoid selection bias and ensure that the sample reflects the population's variability.

Step 4. Sampling Equation: The selection probability for each record is expressed in Equation (6), which defines the sample size in relation to the total population:

$$N = \frac{Number\ of\ selected\ elements}{\text{Number of elements in the population}} \tag{6}$$

In our case, the sample size (n) is 1,000,000, representing approximately 0.115% of the total population. This sample size is suitable for statistical analysis as, according to the Central Limit Theorem, a sufficiently large sample allows the distribution of the sample mean to approximate a normal distribution, facilitating statistical inference.

Step 5 Variance and Sampling Error: To address bias, we considered the sample variance. The population variance (σ^2) was estimated from the sampled data, and the standard error of the mean (E) was calculated using Equation (7):

$$E = \frac{\sigma}{\sqrt{n}} \tag{7}$$

where σ is the population standard deviation, and n is the sample size. The standard error of the mean was 0.00241, suggesting that the mean calculated from our sample is accurate and reliable for estimating the population mean.

Table 4 presents the composition of the obtained sample, which is used in the experiment to develop a mathematical model representing the index for evaluating vehicular traffic performance.

Table 4. Description detailing the number of records and the number of variables in the sample used.

Dataset	Rows	Number of Variables
Muestra_tráfico-incidencias-POIs	1,000,000	35

3.3.3. Stage 3: Development of the Index for Assessing the Performance Level of Vehicular Traffic on Urban Streets

The development of the index to evaluate the level of performance of vehicular traffic on urban streets involves creating a mathematical model that establishes a relationship between a dependent variable, specifically the jam factor (jamFactor) described in Table 2, and a set of independent variables that influence this factor. These independent variables encompass all those mentioned in the same table. The relationship between them is determined through the application of Genetic Programming (GP), a technique that falls within the framework of Evolutionary Computing.

According to Koza's work [31], the Genetic Programming paradigm offers an approach to exploring the space of a set of symbolic expressions, seeking a highly fit individual

expression to solve or approximate the solution to a specific problem. In this context, populations of symbolic expressions are generated through a genetic process based on the Darwinian principle of survival of the fittest. Additionally, an appropriate genetic crossover operator (sexual recombination) is used to mate symbolic expressions, as illustrated in Figure 2 and Algorithm 1 described in Langdon's work [32].

Algorithm 1 Abstract GP algorithm

1. Randomly create an initial population of symbolic expressions from the available primitives
2. **Repeat**
3. Execute each symbolic expression and ascertain its fitness.
4. Select one or two symbolic expression(s) from the population with a probability based on fitness to participate in genetic operations.
5. Create new individual symbolic expression (s) by applying genetic operations with specified probabilities.
6. **Until** an acceptable solution is found or some other stopping condition is met (for example, reaching a maximum number of generations).
7. **Return** the best-so-far individual.

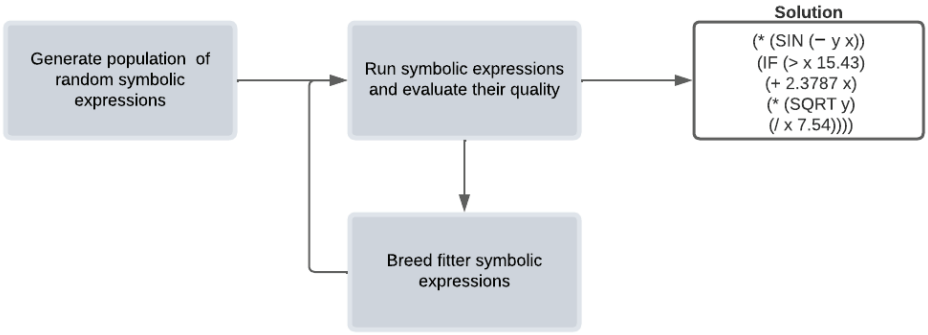


Figure 2. Abstract evolutionary cycle to identify the most suitable symbolic expression, as described in Riccardo Poli's work [33].

The parameters that are needed as input for Algorithm 1, which allow the generation of symbolic expressions relating independent variables to the dependent variable, are described in the following steps:

- Step 1. Identification of the set of terminals (variables and constants): Terminals are fundamental elements used in the construction of symbolic expressions, which aim to solve the posed problem.
- Step 2. Identification of the set of primitive functions: Primitive functions are operations performed on the data, such as COS, SIN, AND, OR, NOT, ADD, and SUB, and will be available to construct the symbolic expressions addressing the problem.
- Step 3. Identification of the reference variable: This is the variable that defines the behavior of the problem at hand.
- Step 4. Identification of the fitness measure: This refers to the fitness function used to evaluate the quality of an individual symbolic expression within the population. Examples of these functions include Mean Squared Error and the Coefficient of Determination, among others.
- Step 5. Assignment of values to the hyperparameters of the evolutionary program: This includes setting the population size, number of generations, crossover probability, and mutation probability, among others.
- Step 6. Implementation.

The development of the index that evaluates the level of performance of traffic in urban streets is carried out using Algorithm 1. This algorithm requires a series of parameters, as detailed in the previous steps, to obtain symbolic expressions. The steps are reiterated below, now incorporating the values from the latest experiment conducted, which led to

the derivation of the symbolic expression that represents the mathematical model of the index for evaluating the performance of vehicular traffic in urban streets.

- Step 1. Identification of the set of terminals (variables and constants): The set of terminals is defined based on the variables listed in Table 2, which refer to traffic, traffic incidents, and points of interest. These variables interact with each other to reflect the state of traffic congestion.
- Step 2. Identification of a sufficient set of primitive functions: Considering the variability in traffic congestion behavior, as illustrated in Figure 3, the following primitive functions are selected: add, sub, mul, div, sqrt, log, abs, neg, inv, sin, cos, and tan.
- Step 3. Identification of the reference variable: An entropy analysis of the 32 variables described in Table 2 is conducted to determine which variable reflects the degree of disorder in vehicular traffic, considering traffic as a system. This analysis identified the variable "jam factor" (jamFactor) as the one that best represents congestion behavior. The jam factor, measured on a scale from 0 to 10 according to Here Maps [34], indicates low vehicular traffic flow (little congestion) with values close to 0, implying a high speed of movement. In contrast, a value close to 10 indicates high traffic flow, reflecting a reduced speed of movement (high congestion). Therefore, the jam factor is inversely proportional to speed, as illustrated in Figure 3 and detailed in Table 5.

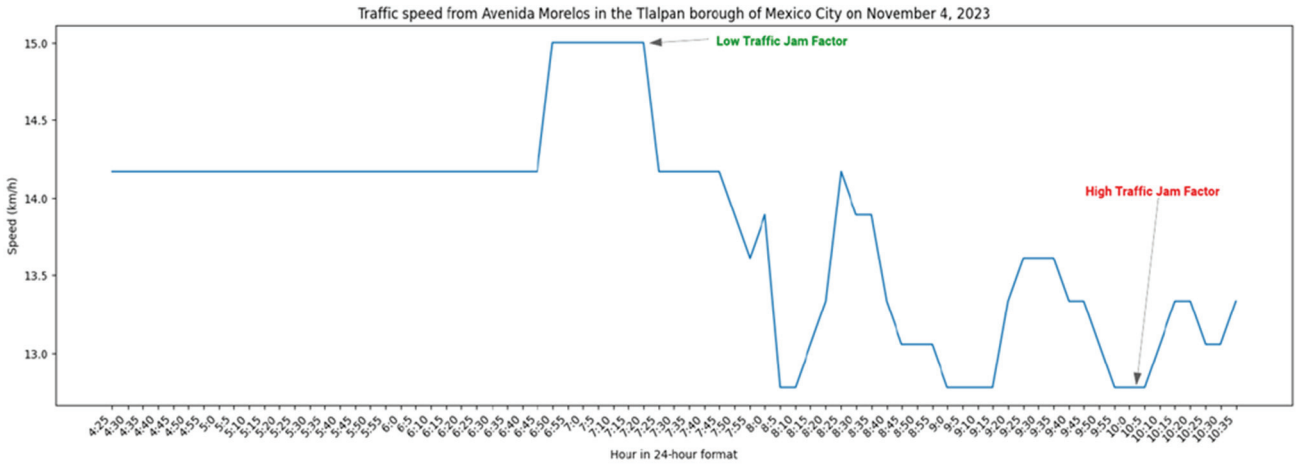


Figure 3. Fragment of vehicular traffic records throughout a day on Avenida Morelos, located in the Tlalpan Municipality.

Table 5. Scale of values for the jam factor developed by Here Maps [34], along with the performance criteria for vehicular traffic proposed in this research.

Range of Congestion Factor	Traffic Congestion Level	Color	Performance Level
0 <= jamFactor < 4	Light congestion	Green	Excellent
4 <= jamFactor < 8	Moderate congestion	Yellow	Good
8 <= jamFactor < 10	Severe congestion	Red	Regular
jamFactor = 10	Road closed	Black	Bad

The performance of traffic congestion, known as the jam factor, reflects the level of vehicular traffic performance on urban streets related with the available speed. This information is detailed in Table 5 and visually illustrated in Figure 4.

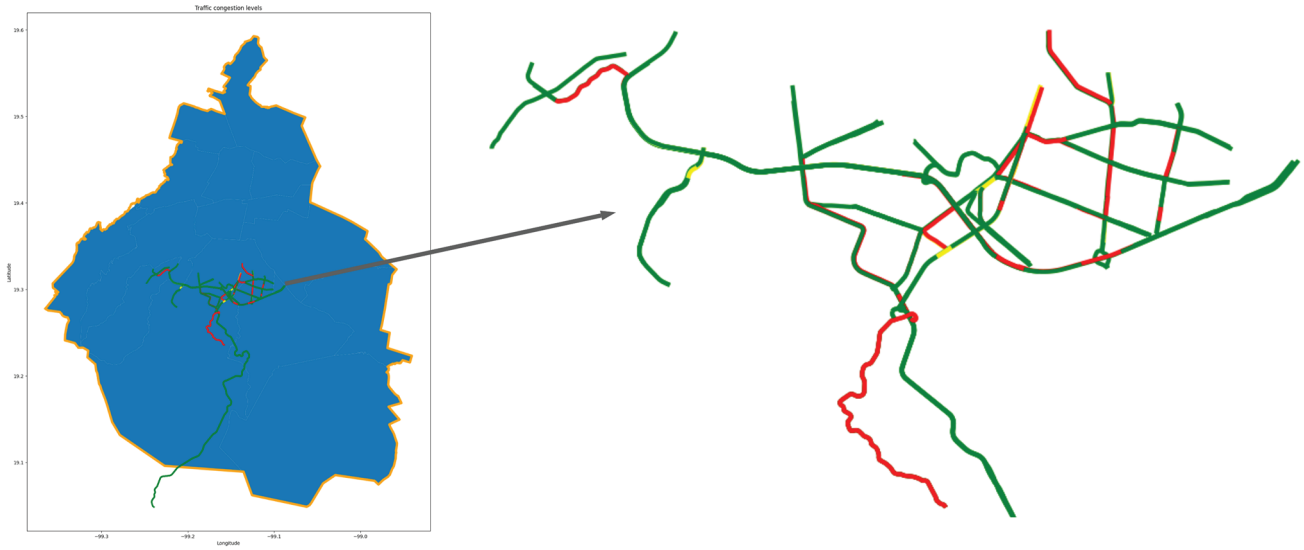


Figure 4. Traffic levels in Tlalpan, Mexico City, using colors: green for light congestion (excellent), yellow for moderate (good), red for severe (fair), and black for road closed (poor), according to the jamFactor.

- Step 4. To select the symbolic expressions that optimally reflect the index evaluating vehicular traffic performance in urban streets, those with the highest fitness are chosen. The fitness functions of Mean Squared Error (MSE) and Mean Absolute Error (MAE) were used to evaluate the most prominent symbolic expressions in each generation.
- Step 5. The selection of hyperparameters in the Genetic Programming (GP) process was carefully considered to maximize the model's ability to find a symbolic expression that aligns with the behavior of the jamFactor, a key indicator in traffic evaluation. The number of generations, set at 300, was chosen to balance solution space exploration while preventing overfitting. With this value, the model has sufficient iterations to improve solutions without risking noise memorization in the data. Additionally, selecting a population of 1 million individuals was essential to ensure a broad and diverse search within the space of symbolic expressions, which is crucial in complex problems like traffic modeling. A large initial set increases the likelihood of the algorithm discovering meaningful relationships among variables. Similarly, the number of GP runs, set at 30, ensures robustness in the results by minimizing the impact of inherent random variability in evolutionary algorithms. This approach ensures that the solutions found are not dependent on a single isolated run but are consistent and repeatable. Finally, the tournament size of 2000 participants was selected to ensure that the selection process within the algorithm favored the exploitation of the best solutions while retaining the ability to explore alternative options in the search space. Collectively, these hyperparameters, shown in Table 6, were designed to achieve an optimal balance between exploring the possible solution space and converging toward a model that accurately reflects the relationship between the jamFactor and the involved variables, thus providing a robust and mathematically valid result for traffic modeling.

Table 6. Hyperparameters used in the evolutionary program.

Parameter	Value
Dataset	Muestra_tráfico-incidencias-POIs
Number of generations	300
Population	1,000,000
Number of executions	30
Tournament size	2000

- Step 6. Implementation: To implement Algorithm 1 using the data defined in Steps 1 to 5, the SymbolicRegressor class from the Gplearn library in Python was utilized. This tool enabled the execution of genetic programming and the generation of symbolic expressions that represent the mobility index.

Once the parameters in Algorithm 1 were configured, the experiment was conducted to obtain a set of symbolic expressions, from which the most suitable one for evaluating the performance of vehicular traffic on urban streets was selected. The experiment involved running Algorithm 1 a total of 30 times, and in each execution, the most optimal symbolic expression was chosen based on its fitness value. Tables 7 and 8 present the symbolic expressions found during the experiment.

Table 7. Symbolic expressions representing the index for evaluating the performance of vehicular traffic on urban streets, obtained in the experiment—Part 1.

Number	Symbolic Expression	MSE	MAE
1	1.26275464756236*sqrt(0.627136457955492*Abs(log(−(X14 − Abs(log(−X5) − Abs(log(Abs(X5))−Abs(X5)))))−Abs(X5)) + 0.627136457955492*Abs(X14 − log(−(X14 − Abs(log(−X5) + log(Abs(sqrt(−X5))) − log(Abs(log(X5)))))) + Abs(X5)) + 1)	9.290168	1.921352
2	2*X16	5.60721	1.8111
3	Abs(log(X6*log(log(Abs(log(log(log(X6)**(1/4)) − log(log(Abs(log(−log(log(log(log(X6)))/(X8 + log(log(log(X6)))/(X8 + sqrt(log(log(X6))/Abs(sqrt(log(log(log(log(log(log(X6)))/(X8 + log(X6) + log(log(log(X6)))))))/(X8 + log(log(X6))/log(log(log(X6))))) − log(log(X6))))))) + log(log(X6)**(1/4)))))**(1/8)))))**(1/4))/(X11 + log(log(log(X6)**(1/4)))))))	9.290168	1.921352
4	log(X2 + log(log(X17) + log(X5 + log(X17) + log(X2 + log(X17) − log(log(X17)))) + zoo)) + log(Abs(X5)) + tan(X2)	9.290168	1.921352
5	sqrt(log(−4*X26 + 3*sqrt(X31) + sqrt(log(−X26 + sqrt(log(-X26 + sqrt(X31) + log(−X26 + sqrt(X31))))))))	9.290168	1.921352
6	Abs(log(log(X30)))	9.290168	1.921352
7	sqrt(−X17 + cos(sqrt(4*X7 + log(X7))) + cos(sqrt(−2*X17 + 5*X7 + cos(log(X7)))) + cos(sqrt(X7 + X20 + log(X7))) + cos(sqrt(2*X7 + log(X7) + cos(X7 + cos(2*X7 + cos(log(X7)))))) + cos(sqrt(2*X7 + log(X7) + cos(X7 + cos(log(X7))))) + cos(sqrt(2*X7 + log(X7) + cos(sqrt(−X17 + sqrt(−X17 + log(X7)) + log(X7))))) + cos(sqrt(2*X7 + 2*log(X7) + cos(sqrt(−X17 + log(X7))))) + cos(sqrt(2*X7 + 2*log(X7) + cos(X7) + cos(log(X7)))) + cos(sqrt(5*X7 + cos(X7 + cos(2*X7 + X20)) + sqrt(cos(log(X7))) + 2*cos(log(X7)))))	9.290168	1.921352
8	X2*(X2/cos(X29*cos(X2*(log(tan(X11)) + tan(X11) + cos(1/cos(X2*X29*(cos(1/X2) + X2/X31)) + 2/X2)/X6))) + cos(1/X2)) + cos(X2/X31)	7.006806	1.716332
9	Abs(sqrt(X17))	9.290168	1.921352
10	X11/X7	3.442246	1.266242
11	log(log(X35*X5) − log(X5 − sqrt(log(X35*X5 + X5 − Abs(sqrt(−X5 + log(X35*X5) − log(X5 − sqrt(log(X35*X5 + X5 − log(sqrt(log(−X35*X5 + X5)))))))*sqrt(log(X28 − X5 + log(X35*X5)))*Abs(sqrt(−X28 + X5 − X5)))))))*sqrt(log(X28 − X5 + log(X35*Abs(sqrt(X28)))))*Abs(sqrt(−X28 + X5 − X5)))	9.290168	1.921352
12	sqrt(tan(tan(sqrt(tan(cos(tan(sqrt(tan(tan(tan(sqrt(tan(tan(cos(cos(tan(1/X4)/X4)/X4)))))**(1/4)))))/X4))))**(1/4)))	9.290168	1.921352
13	sqrt(sin(X10/X3) + 1/log(X5)) + log(X5) + sin(X10/X3)	34.14652	5.041829
14	sqrt(X15)	5.411423	1.737491
15	sqrt(sqrt(X23) + sin(X23**(1/4) + sqrt(X23)) + sin(sqrt(sqrt(sin(X23)) + sin(X23)) + sin(X23)) + sin(sin(X23) + sqrt(sin(sqrt(sin(X23**(1/8) + sin(X23)))))))	8.142053	1.866521

In the Table 7, the ** operator represents exponentiation, and the * symbol represents multiplication.

Table 8. Symbolic expressions representing the index for evaluating the performance of vehicular traffic on urban streets, obtained in the experiment—Part 2.

Number	Symbolic Expression	MSE	MAE
16	sqrt(X5 + sqrt(6*X5 + sqrt(X15 + sqrt(X5) + 15*X5 + sqrt(X15 + 7*X5 + 2*sqrt(log(X15))) + 2*sqrt(log(X15)) + log(X15)) + 2*sqrt(log(X15)))*sqrt(log(X15)))	4.788955	1.60574
17	X11/X7	3.442246	1.266242
18	X11**(1/4)	5.630526	1.771934
19	sqrt(X17)	5.62155	1.73557
20	sqrt(Abs(X15))	9.290168	1.921352
21	log(X11*(2 *tan(X2) + tan(X2) /log(cos(X11)) + 1/(X11*log(log(cos(X11))/(X11*sqrt(X2 + 1/tan(X2))*(X6*tan(X2) + tan(X2))))))*cos(1/tan(X2))/sqrt(Abs(X6)))	9.290168	1.921352
22	X30/X6	3.390861	1.304375
23	X3 − X6	2.621522	1.089678
24	sqrt(log(X4*(X13*X30**(3/2)*X4 *(X13*sqrt(X13*X30)*(−X30 + X7**4) − X30) + X30) + X7 + sqrt(X13*X30)))	0.980081	0.580636
25	X14 + X16	5.562126	1.713949
26	sqrt(X15)	5.411423	1.737491
27	X11/X6	3.472604	1.277578
28	log(log(X15) + Abs(2*log (Abs(X2)) + Abs(X2) + Abs(log(Abs(X2)) + Abs(log(X2) + log(log(Abs(X2) + Abs(log(X5))) + log(Abs(X5)) + Abs(X2)))))) + Abs(log(X5) + Abs(X2))	9.290168	1.921352
29	X14 + tan(X2)	5.32177	1.719235
30	X10**(1/4)	5.413556	1.721787

In the Table 8, the ** operator represents exponentiation, and the * symbol represents multiplication.

The symbolic expression identified in the experiment as the best representation of the index for evaluating the performance of vehicular traffic on urban streets was selected based on its lowest Mean Squared Error (MSE) value. A low MSE value is preferable, as it indicates that the model's predictions are closer to the actual values in the dataset. In this case, the chosen symbolic expression is number 24, detailed in Table 8, which has an MSE of 0.980081. Therefore, the index evaluating the performance of vehicular traffic on urban streets is presented in Equation (8).

$$\text{sqrt(log(X4*(X13*X30**(3/2)*X4 *(X13*sqrt(X13*X30)*(−X30 + X7**4) − X30) + X30) + X7 + sqrt(X13*X30)))} \quad (8)$$

where

X4 = Length of the segment,
X7 = The speed may exceed the legal speed limit,
X12 = Day number of traffic monitoring (1 = Monday, 2 = Tuesday, . . . 7 = Sunday),
X30 = Time block in which the incident was recorded.

The interpretation of the mobility index, expressed in Equation (8), is indeed complex; therefore, a simplification process is carried out, which is described below for Equation (8).

$$\sqrt{\log\left(X_4 \cdot \left(X_{13} \cdot X_{30}^{3/2}\right) \cdot X_4 \cdot \left(X_{13} \cdot \sqrt{X_{13} \cdot X_{30}} \cdot \left(-X_{30} + X_{30}^4\right) - X_{30}\right) + X_{30}\right) + X_7 + \sqrt{X_{13} \cdot X_{30}}} \quad (8)$$

Step 1 Using the properties of logarithms, separate the terms in Equation (8) to obtain the following expression:

$$\log(X_4) + \log\left(X_{13} \cdot X_{30}^{3/2}\right) + \log\left(X_4 \cdot \left(X_{13} \cdot \sqrt{X_{13} \cdot X_{30}\left(-X_{30} + X_{30}^4\right) - X_{30}}\right)\right) + \log(X_{30}) + \log(X_7) + \log\left(\sqrt{X_{13} \cdot X_{30}}\right)$$

Step 2 Reorganize and group the terms related to X_4 and X_7 to obtain the following expression:

$$\sqrt{log(X_4) + \alpha_1 + \beta_1 \cdot log(X_7) - \beta_2 \cdot log(X_{30}) + \gamma}$$

where α_1 and β represent combinations of the other terms.

Step 3 Substitute logarithms, separating $log(X_7)$ from X_{30}, resulting in the following expression:

$$\sqrt{log(X_4) + \beta \cdot log\left(\frac{X_7}{X_{30}}\right) + \gamma}$$

Step 4 Implement the series approximation to finally transform Equation (8) into Equation (9), as shown below:

$$\alpha \cdot X_4 + \frac{\beta \cdot X_7}{X_{30}} + \gamma \tag{9}$$

Equation (9) results from processing the original Equation (8) in the proposed model for evaluating vehicular traffic performance on urban streets. This simplification not only optimizes the mathematical complexity of the model but also provides a clear and accessible interpretation of how different variables impact traffic performance, which is crucial for traffic planners.

First, the term $\alpha \cdot X_4$ represents the direct influence of the variable X_4, which refers to the segment length associated with vehicle density on a road or even specific characteristics of the road type (such as capacity). The coefficient α provides a quantitative indicator of how variations in X_4 affect traffic performance. For example, if vehicle volume on a street increases, overall traffic performance is expected to decrease, as higher density tends to cause congestion. This relationship allows planners to anticipate potential congestion issues and adjust their strategies, such as implementing adaptive traffic signals or modifying recommended routes to optimize traffic flow.

The second term, $\frac{\beta \cdot X_7}{X_{30}}$, introduces a relationship between X_7, which represents speeds exceeding the legal limit and may correlate with the number of traffic incidents or accidents, and X_{30}, which refers to the time block in which incidents occur. Here, β acts as a multiplier that quantifies the influence of incidents on traffic performance, taking road capacity into account. This term indicates that the impact of traffic incidents is more significant when road capacity is limited. Therefore, if there is an increase in X_7, there is a higher likelihood of incidents, which will result in greater congestion, especially on already saturated roads (under X_{30}). This insight is essential for planners, as it enables them to prioritize interventions in accident-prone areas or implement more effective prevention and response measures.

Finally, the constant term γ can encompass additional factors affecting traffic performance, such as weather conditions, roadwork, or special events that are not directly measured by X_4 or X_7, such as traffic days. Including γ in the model allows planners to account for external variations that can influence traffic and are equally relevant to managing road infrastructure. By integrating these factors into planning, more robust strategies can be developed that consider not only typical traffic conditions but also atypical situations that may arise.

In this experiment, the lowest value of the Mean Absolute Error (MAE) also corresponded to symbolic expression number 24, detailed in Table 8, which showed a value of 0.580635, making it the best choice in terms of fitness and value adjustment.

3.3.4. Stage 4: Index for Assessing the Performance Level of Vehicular Traffic on Urban Streets Testing

The evaluation of the index that measures the performance of vehicular traffic in urban streets involves analyzing its effectiveness and ability to identify congestion on these roads. This is achieved by applying the model to an independent test dataset, followed by a comparison between the model's estimates and the actual or expected values.

The validation of the proposed mobility index, represented in Equation (9), was conducted using a new dataset comprising 1152 records corresponding to four days of vehicular flow in the Tlalpan Municipality. The results obtained from the index in Equation (9) were compared with the jamFactor values provided by Here Maps as a reference. Tables 9 and 10 present an excerpt of the results derived from the variables involved in the index proposed in this research.

Table 9. Evaluation of the proposed index compared to the jamFactor value provided by Here Maps—Part 1.

Row	Proposed Index	Performance Level by the Proposed Index	jamFactor	Performance Level by the Here Maps
1	2.877694908	Excellent	0	Light congestión
2	2.974609305	Excellent	3.2	Light congestión
3	3.148829003	Excellent	0	Light congestión
4	2.722841223	Excellent	8.5	Severe congestion
5	2.956925733	Excellent	8.1	Severe congestion
6	2.994402723	Excellent	1.7	Light congestión
7	2.910501036	Excellent	2.5	Light congestión
8	2.884211067	Excellent	1.7	Light congestión
9	3.147885353	Excellent	4.4	Moderate congestion
10	3.671126049	Excellent	0	Light congestión
11	3.661933767	Excellent	0	Light congestión
12	3.175902725	Excellent	2.5	Light congestión
13	3.152494331	Excellent	2.5	Light congestión
14	3.331350929	Excellent	8.1	Severe congestion

Table 10. Evaluation of the proposed index compared to the jamFactor value provided by Here Maps—Part 2.

Row	Proposed Index	Performance Level by the Proposed Index	jamFactor	Performance Level by the Here Maps
15	3.049161244	Excellent	8	Severe congestion
16	3.644403954	Excellent	0	Light congestión
17	2.817223	Excellent	9.3	Severe congestion
18	3.013098766	Excellent	1.1	Light congestión
19	2.950489174	Excellent	2.6	Light congestión
20	3.130873468	Excellent	0	Light congestión
21	2.973544013	Excellent	3.1	Light congestión
22	2.965452564	Excellent	8.7	Severe congestion
23	3.173293495	Excellent	0.4	Light congestión
24	2.934459125	Excellent	7.1	Moderate congestion
25	3.189306155	Excellent	0.9	Light congestión
26	2.956048954	Excellent	3.1	Light congestión
27	2.994039893	Excellent	2.3	Light congestión
28	3.143840567	Excellent	1.4	Light congestión

The accuracy of the proposed index for evaluating vehicular traffic flow on urban streets was determined through the percentage representation of the Mean Squared Error using Equation (10).

$$MSE\% = \frac{MSE}{(b-a)^2} \cdot 100 \quad (10)$$

where

a The minimum value of the observed data set,

b The maximum value of the observed data set.

An MSE of 30% was obtained, indicating that the model has an approximate deviation of 30% from the observed values. This result is significant, as it suggests that the model performs reasonably well in capturing the dynamics of vehicular traffic, adequately reflecting the overall trends in the data. Although there is room for improvement, this level of accuracy provides a solid foundation for the evaluation and management of traffic in urban streets, which can facilitate the implementation of effective strategies to mitigate congestion.

3.3.5. Stage 5: Comparison of Index for Assessing the Performance Level of Vehicular Traffic on Urban Streets with Another Model

The comparison of the traffic index proposed in this research with the INRIX congestion index is both relevant and necessary for various reasons, from theoretical to practical perspectives, in evaluating urban traffic performance.

The INRIX congestion index is one of the most widely used and recognized in the field of traffic analysis. Over the years, it has been validated in multiple studies and applied in various cities around the world, becoming a benchmark standard in traffic performance evaluation [35]. By contrasting our index with that of INRIX, not only is a solid reference framework established, but the effectiveness and applicability of the new model are also evaluated in comparison with a consolidated and prestigious approach.

This comparison provides the opportunity to evaluate the accuracy and robustness of the proposed model. By using standard metrics such as Mean Squared Error (MSE) and Mean Absolute Error (MAE), it is possible to quantify the differences between the indices, offering an objective view of how our index behaves in relation to an established model [36]. This type of analysis is crucial for identifying areas for improvement and validating whether our index can become a viable tool for traffic management.

Moreover, by comparing the results of our index with those of INRIX, we can better understand how each model responds to different conditions of the related variables. This not only helps place our index within the broader context of indicators but also provides valuable insights for developing more effective strategies for traffic management [20,37].

The comparison also allows us to identify limitations in the INRIX index that can be addressed by the new model. By analyzing how each index reacts to specific traffic situations, opportunities can be discovered to improve the proposed index, whether by incorporating new variables, adjusting its methodology, or adopting hybrid approaches that integrate elements from both models. Finally, by contrasting our index with that of INRIX, we contribute to the existing body of knowledge on urban traffic evaluation.

To compare the index obtained in Section 3.3.3, which evaluates the performance of vehicular traffic on urban streets, we contrast it with a model proposed by INRIX in the literature (see Table 11). The INRIX traffic congestion index establishes a relationship between the actual average travel speed and the free-flow speed. This comparison allows us to analyze the effectiveness of the proposed index in relation to a recognized model in the field of traffic analysis.

Table 11. Vehicle congestion models documented in the literature.

Índex	Mathematical Model	Variables	Author
Traffic congestion INRIX	(V_average/V_free)—1.	V_average = Average speed V_free = Free-flow speed	[37]

The comparative analysis between the index proposed in this research and the INRIX index was conducted using a dataset of 2000 records, along with the jamFactor values provided by Here Maps. To evaluate the accuracy of each model, metrics such as Mean Squared Error (MSE) and Mean Absolute Error (MAE) were employed. The results of this accuracy evaluation are presented in Table 12.

Table 12. Evaluation of model performance using MSE and MAE.

Metric	Proposed Índex	Traffic Congestion INRIX
MSE	7.05548636	7.64142780
MAE	2.33575241	1.72769129

The proposed index has an MSE of 7.05548636, which is slightly lower than the INRIX congestion-based index, which has an MSE of 7.64142780. This suggests that the proposed index fits the data better according to this metric. However, it is important to note that the MAE for the INRIX index is lower, with a value of 1.72769129, indicating that the former may be slightly more accurate than the proposed index.

In the field of traffic management, it is essential not only to assess the overall accuracy of predictions but also to consider the model's ability to identify trends and patterns at a macro level. The lower MSE obtained in the proposed model suggests that it adjusts more effectively to traffic variations, which is especially valuable for urban planning and decision-making. This research proposes a model that clearly relates the variables involved in analyzing traffic performance on city streets. The precise identification of congestion patterns can facilitate more efficient and optimized management of road resources.

Additionally, the practical applicability of the proposed model can be strengthened by integrating it with other traffic analysis and management systems. Although the MAE is higher, this does not reduce the model's utility in contexts where understanding traffic dynamics and making forecasts based on historical trends is crucial. For example, if the model can anticipate critical congestion moments or traffic patterns that are not easily detectable through other metrics, its implementation could be highly beneficial, thus contributing to more proactive and effective urban traffic management.

4. Results

The analysis was conducted using a dataset of 1,000,000 records and 35 variables related to traffic, services, and incidents on the roads, employing the congestion factor as the reference variable. Through genetic programming, the relationship between these variables was explored with the aim of obtaining a symbolic expression that represents an index to evaluate the performance of vehicle traffic on urban streets. The resulting symbolic expression, presented in Equation (8), reveals a nonlinear relationship between the various variables that comprise the mobility index, where each variable contributes uniquely and specifically to the final value of the index.

For example, the variable X4, which represents the length of the segment, highlights the importance of the street segment size in traffic flow. As the length of a segment increases, traffic tends to be more fluid; in contrast, shorter segments often have multiple intersections, which can lead to higher density. Additionally, the variable X7, which indicates the current or real-time speed, is crucial for assessing the speed of vehicles at a given moment.

The day on which traffic occurs, represented by variable X12, also plays a fundamental role, as traffic behavior varies from Monday to Sunday. Variable X30, which denotes the time block in which vehicle flow is recorded, significantly impacts traffic performance. Together, these variables reflect key aspects of urban traffic that influence congestion. Equation (8) allows for the quantification and evaluation of these factors to estimate the level of congestion in a specific urban environment. Furthermore, mathematical operations such as square root, logarithm, and operators like addition, subtraction, multiplication, and division are incorporated to process some of the variables when calculating the final

value of the index. This indicates that the model takes into account both the magnitude and nature of the variables in its assessment of vehicle traffic performance.

The proposal of the index with an adjustment treatment, represented by Equation (9) in this work, provides a tool that allows for a quantitative estimation of vehicle traffic performance on urban streets. Unlike other indices, the proposed index details the relationship between variables that are not reflected in other models, making it a highly descriptive index that clearly indicates which variables are associated with optimal or poor vehicle traffic. This index can be of great utility for traffic planners, municipal authorities, and others interested in understanding and managing traffic congestion in urban environments. The equation offers an objective and data-driven tool for assessing congestion, which can facilitate decision-making and the design of strategies aimed at improving urban mobility.

The comparison of the proposed index in this study with other indices present in the literature involves evaluating the accuracy of the developed model in relation to existing alternatives. This analysis allows us to determine the relative capacity of our model to assess vehicle traffic performance on urban streets, contrasting it with approaches used in previous research. To this end, we compare the index obtained in Section 3.3.3, which evaluates vehicle traffic performance on urban streets, with a model proposed by INRIX in the literature. The INRIX traffic congestion index establishes a relationship between the actual average travel speed and the free flow speed. This comparison enables us to analyze the effectiveness of the proposed index in relation to a recognized model in the field of traffic analysis.

The comparative analysis between the proposed index in this research and the INRIX index was conducted using a dataset of 2000 records, along with the jamFactor values provided by Here Maps. To assess the accuracy of each model, metrics such as Mean Squared Error (MSE) and Mean Absolute Error (MAE) were employed. The results of this accuracy evaluation indicate that the proposed index has an MSE of 7.05548636, which is slightly lower than the INRIX-based congestion index, which presents an MSE of 7.64142780. This suggests that the proposed index fits the data better according to this metric. However, it is important to note that the MAE for the INRIX index is lower, with a value of 1.72769129, indicating that the latter may be slightly more accurate than the proposed index.

Although the proposed index is superior to INRIX in MSE and INRIX is more accurate in terms of MAE, the proposed index offers significant advantages regarding the understanding of traffic performance on the streets. Unlike INRIX, which simply classifies the performance of a street on a scale, the proposed index not only evaluates whether traffic is good or bad but also identifies and relates various factors that influence these dynamics. This allows users to have a clearer perspective on what is affecting traffic, providing a greater expectation of understanding regarding the current conditions and the reasons behind the observed performance.

5. Discussion

The evaluation of traffic congestion in urban environments constitutes a complex challenge that requires consideration of multiple interrelated factors. In this study, a model referred to as the "index" is proposed, developed through genetic programming, to assess the performance of vehicular traffic on urban streets. Our research represents a significant contribution to addressing this issue, although it also raises various considerations that require deeper analysis and discussion.

The comparison of the results predicted by the proposed model with other types of traffic data, as well as evaluating its performance in different urban contexts and traffic conditions, would be a valuable approach to strengthening the model. A more thorough exploration of the influence of the selected variables in the mobility index equation may require additional manual adjustments. This would involve incorporating new variables empirically known to impact traffic congestion. Nevertheless, it will be necessary to reevaluate the model to determine its effectiveness with the newly added variables and to verify whether these improve its performance.

It is essential to detail how the index model could be applied in practice to address the evaluation of traffic congestion in urban environments. This involves exploring its potential use in specific urban planning and traffic management applications, as well as identifying the challenges and limitations that may arise during its implementation.

6. Conclusions

This study presents a comprehensive perspective on assessing the performance of vehicle traffic on urban streets, using an index derived through genetic programming and symbolic regression. The application of this model and the analysis of its results have led to significant conclusions that illuminate the nature and management of traffic congestion in urban areas.

The use of genetic programming to derive a symbolic expression that represents the index relating the most impactful variables of vehicle traffic performance on urban streets proves to be an effective strategy for capturing the complexity of congestion in urban environments. The variables selected in the equation, such as traffic origin, incident severity, expected speed on the road, and the number of segments, reflect important aspects that influence urban traffic congestion.

The evaluation of vehicle traffic performance through the proposed index offers a systematic and precise approach to measuring congestion. This model can be invaluable for informed decision-making in traffic management and urban mobility. Furthermore, the developed index has the potential to be applied in various areas of urban planning and transportation policy formulation, providing an objective, data-driven tool for assessing and addressing congestion in urban environments.

However, it is essential to discuss the broader implications of this research in the context of traffic planning and policy development. Understanding traffic congestion should not be limited to measurement; it must be integrated into a framework that considers the long-term effects on urban mobility and the well-being of citizens.

One of the main implications of this study is that integrating the proposed index into traffic planning can transform how urban congestion is addressed. Instead of simply identifying the most congested streets, urban planners can use the index to understand the underlying factors contributing to congestion. This allows for the development of more tailored solutions, such as modifying routes, optimizing traffic light timings, and redistributing traffic based on actual observed conditions.

Moreover, the index can be useful in prioritizing infrastructure investments. Areas that show poor performance in the index could be candidates for infrastructure improvements, such as lane expansions, the creation of exclusive public transport routes, and the implementation of more sophisticated traffic management systems. This not only contributes to improving traffic flow but can also have a positive impact on reducing pollution and enhancing the quality of life for residents.

In the realm of transportation policies, this study emphasizes the need for a holistic, data-driven approach to policy formulation. Decision makers can use the index to establish key performance indicators that guide traffic policies. This could include creating incentives for the use of sustainable mobility alternatives, such as public transport, cycling, or promoting shared mobility.

Furthermore, by providing accurate and up-to-date data on congestion, the index can be crucial in formulating policies that respond to the changing needs of urban communities. In periods of high demand, such as sports events or festivals, the index could help establish temporary management strategies that minimize the impact of increased traffic.

To further enrich the proposed mobility index model, future research should consider applying the index in different urban contexts and evaluating its effectiveness in various cities with distinct demographic and geographical characteristics. It would also be beneficial to explore the index's ability to adapt to exceptional situations, such as emergencies or natural disasters, where traffic may be significantly affected. Additionally, the comparison of the proposed index in this work is only conducted against the INRIX model; therefore,

strengthening the proposed model would involve including more diverse and relevant benchmarks. Incorporating metrics such as the Speed Performance Index and the Speed Reduction Index could provide a more comprehensive view of traffic performance. The Speed Performance Index measures the effectiveness of the average vehicle speed compared to the ideal speed, providing context on how traffic behaves concerning urban mobility objectives. Meanwhile, the Speed Reduction Index evaluates the decrease in speed compared to a free-flow scenario, allowing for the identification of critical areas where interventions are necessary to improve traffic flow. This would not only enhance the understanding of the relative performance of the proposed index but could also provide valuable insights into its robustness and applicability across a variety of traffic scenarios. Including multiple comparisons could facilitate a more thorough analysis and contribute to the validation of the index in practice, which, in turn, could guide the formulation of more effective traffic policies tailored to the specific realities of each urban context.

Similarly, as future work, a detailed methodology is proposed to integrate the mobility index into current traffic management systems, with the aim of optimizing its utility for urban planners and enhancing its impact on decision-making. This methodology would begin with a comprehensive evaluation of technical compatibility, ensuring that the index can be efficiently integrated into the already implemented traffic management systems. To achieve this, the most relevant variables of the index that provide significant value would be identified, and the desired outputs would be established, such as real-time alerts and congestion reports, that can be adapted to the operational functions of these systems. In the next phase, an integration infrastructure would be developed, likely through APIs and data visualization modules, to enable the real-time flow of index data to traffic systems. These tools would include accuracy and quality tests to ensure the reliability of the information, as well as the creation of visual interfaces that simplify the interpretation of the index in interactive maps and graphs, thereby improving strategic decision-making regarding traffic.

As an additional future work, a detailed methodology is proposed for integrating the mobility index into current traffic management systems, aiming to optimize its utility for urban planners and enhance its impact on decision-making. This methodology would begin with a comprehensive assessment of technical compatibility, ensuring that the index can be efficiently integrated into already implemented traffic management systems. To achieve this, the most relevant variables of the index that provide significant value would be identified, and desired outputs, such as real-time alerts and congestion reports, would be established to adapt to the operational functions of these systems. In the next phase, an integration infrastructure would be developed, likely through APIs and data visualization modules, allowing for the real-time flow of index data into traffic systems. These tools would include accuracy and quality testing to ensure the reliability of the information, as well as the creation of visual interfaces that simplify the interpretation of the index in interactive maps and graphs, thereby improving strategic decision-making regarding traffic.

To implement the proposed index in real-time traffic monitoring systems using technologies such as the Internet of Things (IoT) and Intelligent Transportation Systems (ITS), several computational and operational requirements are necessary. First, it is essential to have adequate processing capacity, including fast processing units (CPU/GPU) and high-performance servers capable of handling large volumes of data in real time. Additionally, an effective storage system is needed, which allows the use of real-time databases and distributed storage solutions to ensure data availability and scalability. Connectivity and bandwidth are equally critical; a robust network infrastructure must be established with sufficient capacity to transmit data from IoT devices to analysis servers, ensuring low latency. Furthermore, interoperability is essential, which implies implementing standardized communication protocols and APIs that facilitate the integration of various data sources. For data analysis, real-time analysis algorithms and visualization tools must be implemented to enable urban planners to easily interpret the information. Security protocols must also be considered to protect data integrity, along with a security monitoring system that alerts to potential threats. The system architecture should be scalable to accommodate

future demands and the integration of new IoT devices, and finally, it is important to have trained personnel to manage and maintain the system, ensuring its proper functioning and updates. These requirements are essential for the proposed index to provide accurate and useful analyses of real-time traffic performance, facilitating informed decision-making to improve urban mobility.

Author Contributions: Conceptualization, E.D.l.C.-N. and A.M.-R.; methodology, E.D.l.C.-N.; software, E.D.l.C.-N.; validation, H.E.-E., A.M.-R. and E.D.l.C.-N.; formal analysis, E.D.l.C.-N.; investigation, O.A.P.-M.; resources, E.D.l.C.-N.; data curation, O.A.P.-M.; writing—original draft preparation, E.D.l.C.-N.; writing—review and editing, H.E.-E.; visualization, O.A.P.-M.; supervision, A.M.-R.; project administration, A.M.-R.; funding acquisition, O.A.P.-M. and E.C. Interpretation of Results, Analysis of the Hyperparameters of the Genetic Algorithm and Verification of Results. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The traffic and road incident dataset is available through the Here Maps API (https://developer.here.com/develop/rest-apis), while service data is obtained via the Geoapify API (https://www.geoapify.com/places-api/).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. The Complexity Science Center. Available online: https://www.c3.unam.mx/boletines/boletin5.html (accessed on 4 January 2024).
2. IDEA. Movilidad Sostenible. Available online: https://www.idae.es/movilidad-sostenible/la-piramide-de-la-movilidad-urbana (accessed on 4 January 2024).
3. National Institute of Statistics and Geography. Available online: https://www.inegi.org.mx/temas/vehiculos/ (accessed on 4 January 2024).
4. Mondal, S.; Gupta, A. Speed distribution for interrupted flow facility under mixed traffic. *Phys. A* **2021**, *570*, 125798. [CrossRef]
5. Sun, C.; Pei, X.; Hao, J.; Wang, Y.; Zhang, Z.; Wong, S.C. Role of road net-work features in the evaluation of incident impacts on urban traffic mobility. *Transp. Res. Part B* **2018**, *117*, 101–116. [CrossRef]
6. Wang, W.; Yuan, Z.; Yang, Y.; Yang, X.; Liu, Y. Factors influencing traffic accident frequencies on urban roads: A spatial panel time-fixed effects error model. *PLoS ONE* **2019**, *14*, e0214539. [CrossRef] [PubMed]
7. Almeida, A.; Brás, S.; Sargento, S.; Oliveira, I. Exploring bus tracking data to characterize urban traffic congestion. *J. Urban Mobil.* **2023**, *4*, 100065. [CrossRef]
8. Nian, G.; Sun, J.; Huang, J. Exploring the effects of urban built environment on road travel speed variability with a spatial panel data model. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 829. [CrossRef]
9. Kushchenko, L.E.; Kushchenko, S.V.; Baranov, Y.N.; Dnistrenko, N.S. Traffic modeling. *IOP Conf. Ser. Mater. Sci. Eng.* **2019**, *698*, 066034. [CrossRef]
10. Sun, J. Mathematical modeling of traffic. *J. Phys.* **2021**, *2012*, 012060. [CrossRef]
11. Mohan Rao, A.; Ramachandra Rao, K. Measuring urban traffic congestion—A review. *Int. J. Traffic Transp. Eng.* **2012**, *2*, 286–305. [CrossRef]
12. He, F.; Yan, X.; Liu, Y.; Ma, L. A traffic congestion assessment method for urban road networks based on speed performance index. *Procedia Eng.* **2016**, *137*, 425–433. [CrossRef]
13. Liponhay, M.; Valenzuela, J.F.; Dorosan, M.; Dailisan, D.; Monterola, C. A dynamic urban mobility index from clustering of vehicle speeds in a tourist-heavy city. *Appl. Sci.* **2023**, *13*, 12763. [CrossRef]
14. Pandey, A.; Sharma, M.; Biswas, S. Concept of heterogeneity index for urban mixed traffic. *Int. J. Transp. Sci. Technol.* **2023**, *12*, 354–372. [CrossRef]
15. Leitner, D.; Meleby, P.; Miao, L. Recent advances in traffic signal performance evaluation. *J. Traffic Transp. Eng.* **2022**, *9*, 507–531. [CrossRef]
16. Seong, J.; Kim, Y.; Goh, H.; Kim, H.; Stanescu, A. Measuring traffic congestion with novel metrics: A case study of six U.S. metropolitan areas. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 130. [CrossRef]
17. Wang, W.-X.; Guo, R.-J.; Yu, J. Research on road traffic congestion index based on comprehensive parameters: Taking Dalian city as an example. *Adv. Mech. Eng.* **2018**, *10*, 168781401878148. [CrossRef]
18. Arirja Gani, F.; Yoshii, T.; Kurauchi, S. The suitable index of flow and density in the mixed traffic. *IOP Conf. Ser. Earth Environ. Sci.* **2017**, *71*, 012015. [CrossRef]
19. Liu, Y.; Liu, C.; Zheng, Z. Traffic congestion and duration prediction model based on regression analysis and survival analysis. *Open J. Bus. Manag.* **2020**, *08*, 943–959. [CrossRef]

20. Chen, L.; Shi, J.; Cheng, M.; Zhu, H.; Sun, L. Characteristics of urban road non-recurrent traffic congestion based on floating car data. In Proceedings of the 4th International Conference on Electronic Information Technology and Computer Engineering 2020, Xiamen, China, 6–8 November 2020; pp. 120–126. [CrossRef]
21. Bian, C.; Yuan, C.; Kuang, W.; Wu, D. Evaluation, classification, and influential factors analysis of traffic congestion in Chinese cities using the online map data. *Math. Probl. Eng.* **2016**, *2016*, 1693729. [CrossRef]
22. Mahona, J.; Mhilu, C.; Kihedu, J.; Bwire, H. Factors contributing to traffic flow congestion in heterogenous traffic conditions. *Int. J. Traffic Transp. Eng.* **2019**, *9*, 238–254. [CrossRef]
23. Yu, J.; Wang, L.; Gong, X. Study on the status evaluation of urban road intersections traffic congestion base on AHP-TOPSIS modal. *Procedia Soc. Behav. Sci.* **2013**, *96*, 609–616. [CrossRef]
24. Gullotta, G.; Loret, E.; Stewart, C.; Sarti, F. Traffic attractors and congestion in the urban context, the case of the city of Rome. *J. Geogr. Inf. Syst.* **2020**, *12*, 545–559. [CrossRef]
25. Rahman, M.M.; Najaf, P.; Fields, M.G.; Thill, J.-C. Traffic congestion and its urban scale factors: Empirical evidence from American urban areas. *Int. J. Sustain. Transp.* **2022**, *16*, 406–421. [CrossRef]
26. Pi, M.; Yeon, H.; Son, H.; Jang, Y. Visual Cause Analytics for Traffic Congestion. *IEEE Trans. Vis. Comput. Graph.* **2019**, *27*, 2186–2201. [CrossRef] [PubMed]
27. Owais, M.M.; Enieb, M.; Abbas, Y.A. Evaluation and Analysis of Urban Passengers Transport Modes Operation Performance Efficiency. *JES J. Eng. Sci.* **2011**, *39*, 283–299. [CrossRef]
28. National Institute of Statistics, Geography, and Informatics. Available online: https://www.inegi.org.mx/contenidos/productos/prod_serv/contenidos/espanol/bvinegi/productos/historicos/920/702825920845/702825920845_1.pdf (accessed on 8 January 2024).
29. De la Cruz-Nicolás, E.; Martínez-Rebollar, A.; Estrada-Esquivel, H.; Pliego-Martínez, O.A. Methodology to obtain traffic data and road incidents through maps applications. In *Smart Cities*; Springer Nature: Cham, Switzerland, 2023; Volume 1938, pp. 3–17. [CrossRef]
30. Owais, M. Deep learning for integrated origin–destination estimation and traffic sensor location problems. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 6501–6513. [CrossRef]
31. Koza, J. Genetic programming as a means for programming computers by natural selection. *Stat. Comput.* **1994**, *4*, 87–112. [CrossRef]
32. Langdon, W.B.; Poli, R.; McPhee, N.F.; Koza, J.R. Genetic Programming: An Introduction and Tutorial, with a Survey of Techniques and Applications. *Stud. Comput. Intell.* **2008**, *115*, 927–1028. [CrossRef]
33. O'Neill, M.; Poli, R.; William, B.; Langdon, N.F. Mcphee: A field guide to genetic programming. Genet. Program. *Evolvable Mach.* **2009**, *10*, 229–230. [CrossRef]
34. Here Maps Developer. Available online: https://www.here.com/docs/bundle/sdk-for-ios-navigate-developer-guide/page/topics/routing.html#find-traffic-along-a-route (accessed on 12 January 2024).
35. Pishue, B. *INRIX Global Traffic Scorecard*; Scorecard Rep.; INRIX: Washington, DC, USA, 2023; Available online: https://inrix.com/scorecard/ (accessed on 4 January 2024).
36. Hyndman, R.J.; Koehler, A.B. Another look at measures of forecast accuracy. *Int. J. Forecast.* **2006**, *22*, 679–688. [CrossRef]
37. INRIX. Available online: https://inrix.com/press-releases/medidor-inrix-del-trafico-congestion-desploma/#:~:text=Para%20cada%20segmento%20de%20carretera,con%20libre%20circulaci%C3%B3n)%20%E2%80%93%201 (accessed on 20 February 2024).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

smart cities

MDPI

Review

How 3D Printing Technology Makes Cities Smarter: A Review, Thematic Analysis, and Perspectives

Lapyote Prasittisopin

Centre of Excellence on Green Tech in Architecture, Faculty of Architecture, Chulalongkorn University, Bangkok 10330, Thailand; lapyote.p@chula.ac.th

Highlights:

What are the main findings?

- Key benefits of 3D printing include reducing construction time and material waste, lowering costs, and enabling the creation of scalable, affordable housing solutions.
- Existing challenges remain in terms of cost, scalability, and the need for interdisciplinary collaboration among engineers, urban planners, and policymakers for smart cities.

What is the implication of the main findings?

- Three-dimensional printing or additive manufacturing (AM) offers potential pathways for sustainable urban development.
- A roadmap for future research and practical applications of 3D printing in smart cities, contributing to the ongoing discourse on sustainable and technologically advanced urban development, is provided.

Abstract: This paper presents a comprehensive review of the transformative impacts of 3D printing technology on smart cities. As cities face rapid urbanization, resource shortages, and environmental degradation, innovative solutions such as additive manufacturing (AM) offer potential pathways for sustainable urban development. By synthesizing 66 publications from 2015 to 2024, the study examines how 3D printing improves urban infrastructure, enhances sustainability, and fosters community engagement in city planning. Key benefits of 3D printing include reducing construction time and material waste, lowering costs, and enabling the creation of scalable, affordable housing solutions. The paper also addresses emerging areas such as the integration of 3D printing with digital twins (DTs), machine learning (ML), and AI to optimize urban infrastructure and predictive maintenance. It highlights the use of smart materials and soft robotics for structural health monitoring (SHM) and repairs. Despite the promising advancements, challenges remain in terms of cost, scalability, and the need for interdisciplinary collaboration among engineers, designers, urban planners, and policymakers. The findings suggest a roadmap for future research and practical applications of 3D printing in smart cities, contributing to the ongoing discourse on sustainable and technologically advanced urban development.

Keywords: 3D printing; smart cities; additive manufacturing; digital twins; machine learning; sustainable development; smart material; structural health monitoring; repair; artificial intelligence; urban

1. Introduction

Rapid urbanization, resource shortages, environmental deterioration, and the need for sustainable development have plagued cities worldwide in recent years. Innovative solutions to improve urban living conditions and solve the inefficiencies of traditional urban planning and building methods are needed as the world's urban population grows to 68% by 2050 [1]. This yields a growing market for smart buildings from USD 97 billion in 2023 to around USD 400 billion in the next 7 years [2]. From this perspective, the ubiquitous use

of 3D printing technology in various industries today has the potential to transform urban development [3]. The concept of mass customization effectively shortens the time between design creation and prototyping, allowing for the production of goods closer to the consumer as needed [4]. This review work examines how 3D printing innovation makes cities smarter by improving infrastructure, resource management, and community participation.

Smart cities use technology to improve quality of life, sustainability, and resilience. They use several digital technologies and data-driven solutions to improve service delivery, resource utilization, and public engagement in decision making. Smart cities use cutting-edge technologies to monitor and control urban processes in real time, including transportation, waste management, energy usage, and public safety. "3D printing" or "additive manufacturing" is a critical breakthrough that can enhance urban infrastructure, reduce waste, achieve zero-net carbon neutrality, and empower communities [5–7], aligning with the 2023 United Nations (UN) Sustainable Development Goals (SDGs) [8].

One of the biggest benefits of 3D printing is resource management, which is crucial to smart city development. Material waste from traditional building processes degrades the environment and raises expenses [9]. Three-dimensional printing uses the additive manufacturing concept to build items layer by layer using just the resources needed. Precision cuts material waste and construction's environmental impact. Ahmed [10] reviewed 3D printing technology in 2023 and asserted that it provided a solution that significantly diminishes building expenses compared to conventional construction methods: it reduced construction time by 25% for equivalent houses; minimized material costs by decreasing waste by 50–70%; cut labor costs by 50–80%; decreased labor accidents and injuries by requiring only a team of 3–5 individuals to complete the structure; eliminated costs associated with human errors during construction; and crucially saved 35–60% of the overall project cost by eliminating false work. Figure 1 presents the key benefits of using 3D printing for construction to reduce various costs during construction.

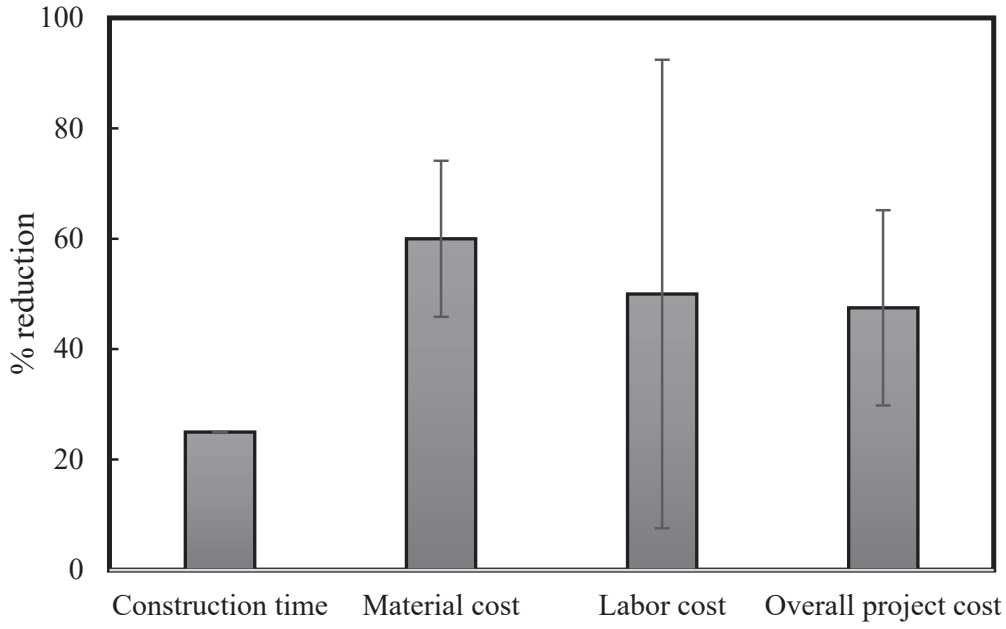


Figure 1. Cost reduction of using 3D printing for construction (source: by author).

Local production with 3D printing reduces transportation, which can boost local economies and reduce carbon emissions. Cities may improve efficiency and community response by creating local 3D printing hubs to make construction components on demand once the solution can be practically implemented [11]. Three-dimensional printing allows quick prototypes and experimentation, which can improve urban infrastructure. Before making adjustments, traditional urban planning requires extensive design procedures and

major infrastructure expenditure. However, 3D printing lets city planners and designers quickly construct scale models and prototypes for iterative design testing and modification [12]. Smart city efforts require adaptation; thus, flexibility is key. Cities can use 3D printing to test public space, transportation, and emergency shelter designs before building them. For example, temporary 3D printed concrete walls and structural self-supported pavilions can be built and fabricated onsite [13–15]. Planners may receive real-time input from citizens and stakeholders to ensure that developments meet community requirements via experimentation.

Housing is another important area where 3D printing can make cities smarter. Many cities have housing shortages, worsening homelessness and affordability. By building affordable homes quickly and cheaply utilizing 3D printing technology, these issues may be addressed. Several projects have shown that 3D printed structures may be practical and attractive. Smart cities can reduce housing instability and improve quality of life by offering scalable, affordable housing alternatives [16]. In addition to infrastructure and housing, 3D printing promotes urban community participation and participatory planning. Cities may guarantee their developments meet community requirements by including individuals in planning and construction. In workshops, tourism, travel, and food-related and community activities, citizens may design public areas, parks, and community buildings using 3D printing [17]. Participants and tourists feel empowered and proud of their urban surroundings with this participatory visualization technique. Strong positive attitude levels of around 64% were obtained when 3D printing was involved in industries in Egypt [18]. Additionally, 3D printing in schools may also raise understanding of urban planning procedures and encourage civic engagement, especially for STEM, social science, and history students [19,20]. Additionally, a notable feature of a garden project in Kenya is the cost. It produced a two-bedroom house at an estimated price of USD 28,000, which is considerably cheaper than the typical price of similar houses by around 35.7%. Furthermore, the initial 10 residences were produced in under 2.5 months [21].

The use of 3D printing in smart cities can help improve sustainability. As cities lessen their environmental effect, 3D printing may help create eco-friendly materials and construction methods. Three-dimensional printing with recycled or bio-based materials supports a circular economy and sustainable development (e.g., with emissions as low as 0.0524 kgCO_2eq/kg) [22,23]. This value is notably low, as compared to the CO_2 emission of ordinary Portland cement of approximately 1 kgCO_2eq/kg. Recycling plastics or organic materials into 3D printing filaments such as recycled polylactic acid (PLA), polyethylene terephthalate (PET), glycol-modified PET (PET-G), ABS, and high-impact polystyrene (HIPS) minimizes waste and encourages sustainable building; thus, research is ongoing [24–27]. Innovative energy-efficient designs like buildings that utilize natural light as well as ventilation can be created using 3D printing. De Rubeis [28] examined the thermal performance of a PLA construction using 3D printing cavities by infrared thermography and a hot box method. His findings indicated that insulating the voids with wool results in thermal uniformity. In addition, Sun et al. [29] conducted an examination of the thermal performance of a 3D printed prototype structure utilizing infrared thermography. The investigation suggested an irregular temperature distribution throughout the cross-section and elevation of the building, signifying disparities in thermal characteristics. The average U-value varied from the U-value at certain sites by as much as 58%.

Since digital technology is expanding rapidly, new innovations have been implemented for cities to become smarter within the contemporary improvement domain for the purpose of improving quality of life [30], including which innovations that are vital for smart cities and those that are beneficial. The literature search revealed the absence of a development summary in existing articles. The main research question of this work is shown below:

Research question: *How does 3D printing make cities smarter?*

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)-based qualitative technique for keywords of screened articles is used to identify relevant characteristics [31], which will be explained in depth.

2. Methodology

The review covers two parts: a qualitative study based on the keywords used and systematic review of existing literature. The systematic review followed PRISMA principles, which have been created for several research domains. The PRISMA process is shown in Figure 2. Studies were retrieved from Scopus, PubMed, and Google Scholar databases. The search terms used were the contexts presented in the titles, abstracts, and keywords of the literature. The terms included "smart city" or "smart building," together with the term "3D printing" or "additive manufacturing." Due to very fast 3D printing development [32], the scope focused on the articles published from 2015–2024. Research from before this decade may be obsolete, and their review may not represent current developments. After the review, it was found that 3D printing in smart cities is underrepresented, thus this article reviews it for smarter cities and smarter buildings. Thus, the study searched Scopus using title, abstract, and keywords in text:

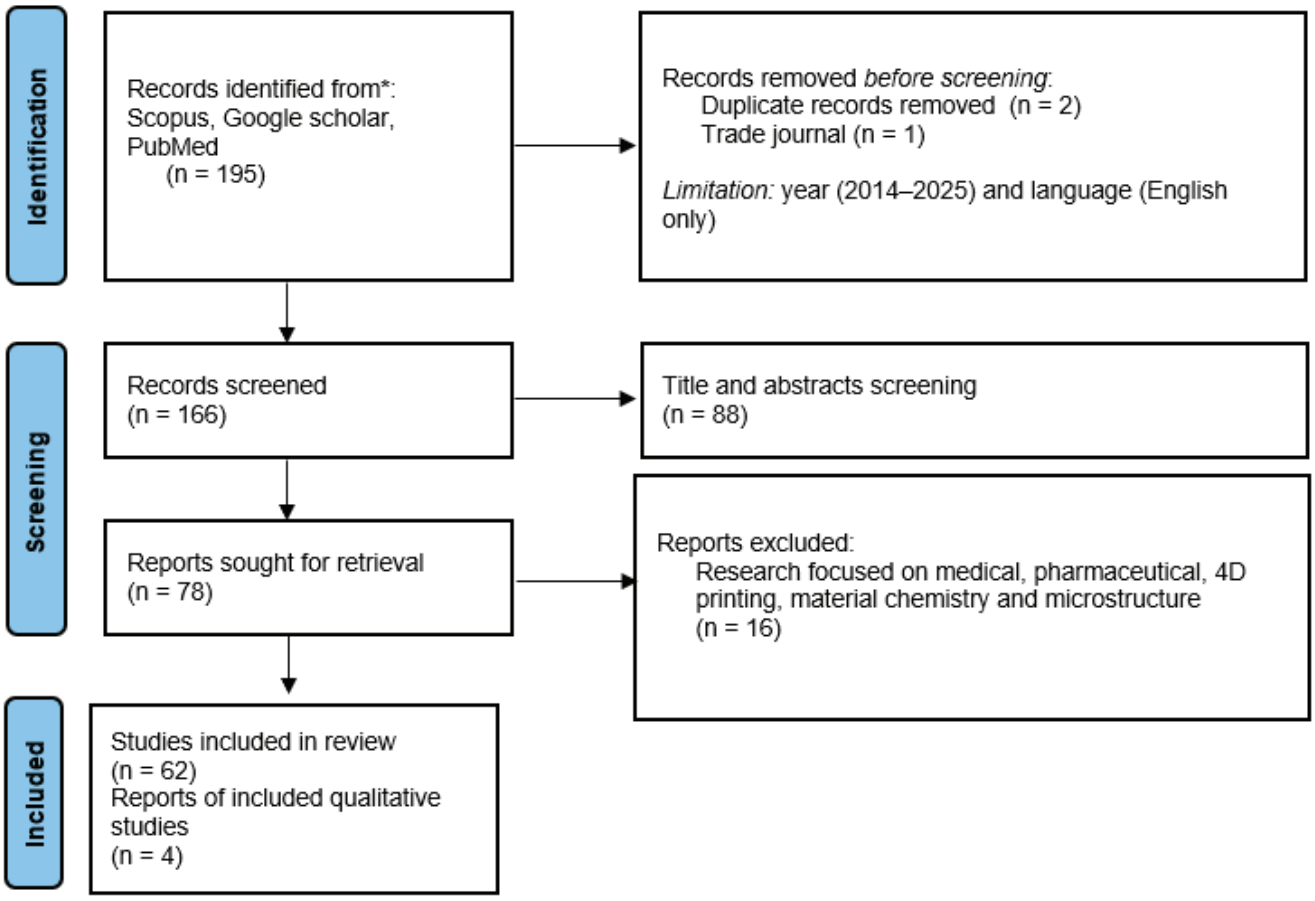


Figure 2. The review approach obtained using the PRISMA guidelines (source: by author). * sources from Scopus, Google scholar, and Pubmed databases.

TITLE-ABS-KEY ((Smart city OR Smart build*) AND (3D printing OR Additive Manufact*)) AND PUBYEAR > 2014 AND PUBYEAR < 2025 AND (LIMIT-TO (LANGUAGE,"English")) AND (LIMIT-TO (SRCTYPE,"j") OR LIMIT-TO (SRCTYPE,"p") OR LIMIT-TO (SRCTYPE,"b")).

2.1. In-/Exclusion Criteria

The present systematic review incorporates the subsequent criteria for inclusion and exclusion.

2.1.1. Inclusion Criteria

- Regarding the utilization of 3D printing from a smart city perspective;
- Including current 3D printing technologies with a published date later than 2015;
- Providing a detailed explanation of the effectiveness that each technology offers;
- Must be in a smart city setting;
- Only English language used.

2.1.2. Exclusion Criteria

- Excluding topics focused on medical, pharmaceutical, automative, 4D printing, material microstructure applications;
- General argument;
- Trade article;
- Based on another study;
- The investigation lacks adequate information regarding the AR/VR technology.

2.2. Data Extraction

During the manuscript evaluation process, the title, abstract, and keywords were examined to verify that the submissions conform to established requirements. This process entails the systematic extraction of several critical core components following a keyword analysis search. After screening, 62 articles were analyzed for this review. The importance of 3D printing was examined. The challenges and prospective research directions are also subsequently addressed for each section. This review study highlights the significance and prospects for academic communities to understand and adhere to the issues raised in this context for future research.

3. Review Analysis and Discussion

3.1. Thematic Analysis on Keywords

Sixty-two pertinent publications were identified from the literature search and screened articles, as given in Table 1. From the thematic analysis, Figure 3 illustrates the number of articles released per year from 2015 to 2024. Analysis indicated that the trend of 3D printing in smart cities is fast escalating. The interests around this issue are growing increasingly pronounced. During the initial stage, extensive research focused on developing smart materials and smart manufacturing. Subsequently, the research advances to AI, machine learning (ML), and digital twin (DT) technology integration. The frequency of terms utilized more than twice was aggregated and is depicted in Figure 4. When ranking the frequency of these keywords from most to least common, the investigation indicates that 3D printing in smart cities, in relation to smart manufacturing, is the most often utilized term. This is succeeded by smart materials, DT, Industry 4.0, soft robotics, AI, SHM, and wire-arc additive manufacturing (WAAM). The keywords were subsequently examined for each segment. The phrases “3D printing,” “additive manufacturing,” and “smart city” were omitted as they are already contextual to the study. Also, AI has been addressed in most other topics, so the discussion of AI/3D printing for smart cities is not mentioned here. After examining 10 important aspects in this thematic analysis, as illustrated in Figure 5, each section is discussed to elaborate on its role in contributing to smart cities. Such technological aspects related to 3D printing can offer a contribution to how the current development of such aspects can benefit the smart city context. The subsequent sections present contributions and significance for the stated aspects.

Table 1. Existing literature determined in this review (source: by author). ✓presents the topic related to the publication.

| Topic | Year | | |

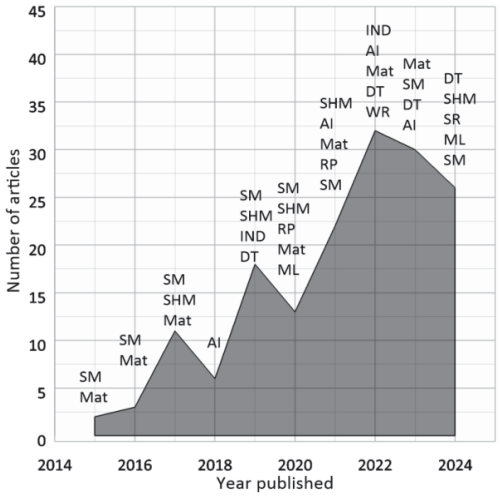


Figure 3. Number of publications published and keywords used each year from 2015–2024 (SM = smart manufacturing; Mat = smart material, SHM = structural health monitoring, Al = artificial intelligence, IND = Industry 4.0, DT = digital twin, RP = repair, ML = machine learning, and WR = wire-arc additive manufacturing) (source: by author).

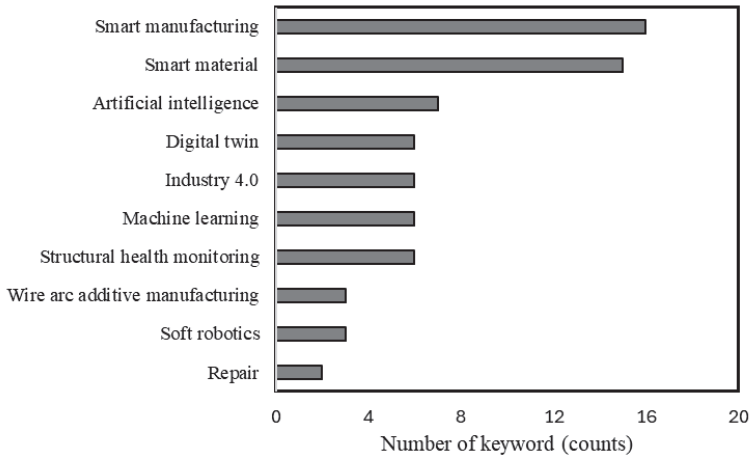


Figure 4. Count of keywords used in this literature review (source: by author).

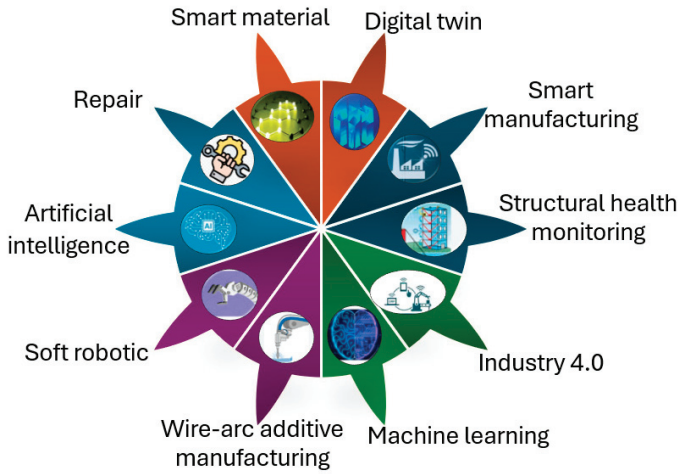


Figure 5. Current technological aspects related to 3D printing for smart city context (source: by author).

3.2. Digital Twins (DTs) in 3D Printing for Smart Cities

DT technology is defined as a virtual representation of physical entities that disclose, forecast, and enhance their tangible counterparts [95]. The term "twin" evokes a simpler framework and implies an equitable relationship between the two "twins." The term "twin" serves as a metaphor, conveying both a relationship of similarity (the digital representation possesses identical properties to the physical thing) and a relationship of parity (the digital representation corresponds to the physical entity as an equal). DTs are delineated into five components, entailing physical, data, analytical, virtual, and connection environments [96].

DT technology is progressively used to model the performance of intricate systems, including those employed in the building and manufacturing industries. Integrating AI enhances the accuracy of these simulations and their capacity to manage extensive data, facilitating predictive maintenance and process optimization. Rojek et al. [37] examined the influence of digital transformation on enhancing 3D printing methodologies using DT virtual representations of physical entities and mentioned that it facilitates real-time surveillance and enhancement of industrial operations. This research emphasizes that DTs serve as a potent instrument for forecasting and alleviating challenges in 3D printing, including flaws and inefficiencies. In terms of construction technology, DTs can be used in interior design and fabrication of architectural and building elements in existing buildings, as depicted in Figure 6. This DT adoption can reduce survey time, errors, and inspectors since the operating process can be performed in a digital file. In addition, DTs can be used in the simulation stage before printing actual buildings. The wall panels, flat slabs, and other structural components can be efficiently printed and fabricated with other construction additives to improve their performance and functionality, such as ultra-high strength (more than 100 MPa), as well as thermal and acoustic insulations [96,97]. For instance, with different printing configurations, the structural-performance-to-weight ratios of 3D printed wall panels can be enhanced more than 200% compared to the original design [98].

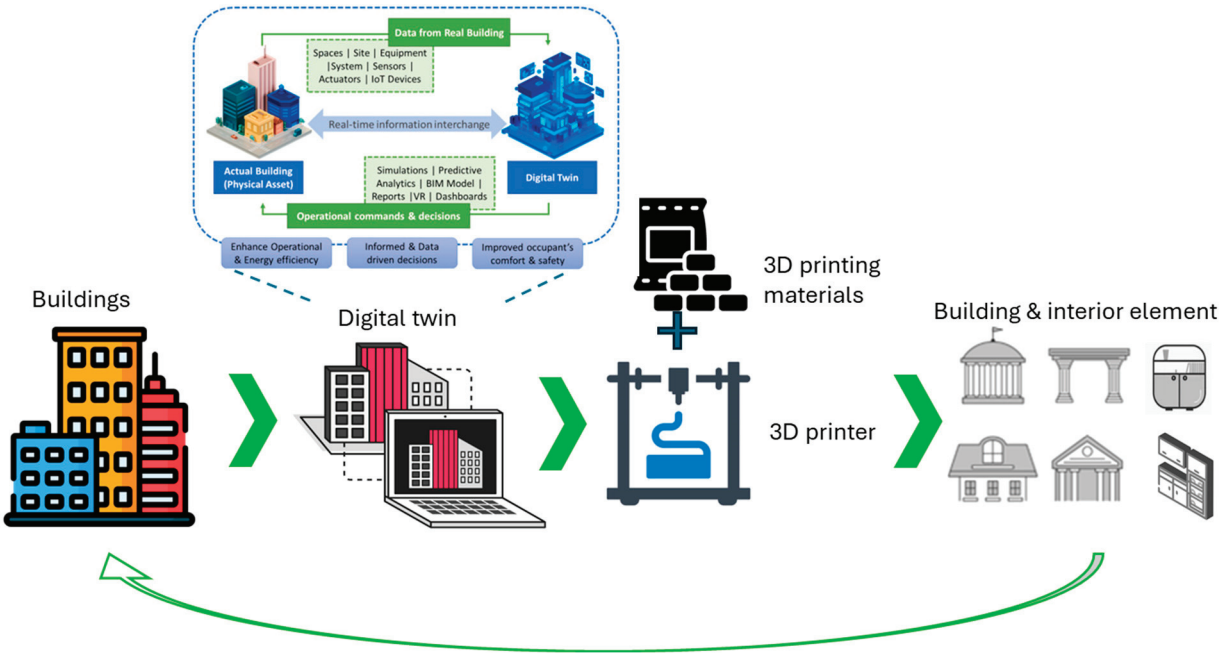


Figure 6. Three-dimensional printing process with DTs, where the DTs can be implemented in the renovation and interior of existing buildings (source: by author).

In the contexts of urban planning for smart cities, DTs generally support 3D city simulations regarding several factors, such as environmental, spatial, historical, and in situ collected data [99,100]. The 3D printing can be modeled and used as a preliminary study from the data obtained from geographical information systems (GISs). This benefits

continuous learning and adaptation, providing more accurate predictions and facilitating better decision-making processes for urban planning, such as categorization of the land use [101]. This approach is particularly valuable in built environments, where the optimization of resources and reduction of downtime are critical for maintaining competitiveness and sustainability [102].

Although the future of DTs in 3D printing appears promising, there remain obstacles to their broad adoption. DTs necessitate considerable initial investment in technology and training, potentially posing a barrier for smaller enterprises. Furthermore, the integration of digital twins with 3D printing necessitates defined data protocols and enhanced compatibility among software systems [37]. As smart city projects gain traction globally, it is imperative for governments, international organizations, and significant businesses to advocate for these standards, therefore enabling more seamless integration [103–105]. DTs can be augmented with several technologies for smart cities such as virtual reality, geographic information systems, life cycle assessment, parametric design, generative design, and building/city information models [95,106–112]. Indeed, 3D printing with DTs also offers another promising technology as a smart tool for mass-customized manufacturing.

3.3. Three-Dimensional Printing in Industry 4.0 for Smart Cities

Industry 4.0, commonly known as the Fourth Industrial Revolution, signifies a radical change in industrial operations via the integration of modern digital technology in manufacturing and other sectors. Industry 4.0 is defined by the integration of physical and digital realms, utilizing technologies such as the Internet of Things (IoT), AI, cyberphysical systems, cloud computing, and big data to establish highly automated, efficient, and linked industrial ecosystems. As Industry 4.0 progresses, its impact on industries such as building, manufacturing, and urban development—especially in smart cities—along with the mentioned technologies grows more substantial [113–116].

Three-dimensional printing revolutionizes built environments within the framework of Industry 4.0 in the era of "fab cities" and "maker cities" [117]. It facilitates the fabrication of intricate structures with exceptional precision and minimum waste through the methodical stacking of materials. It enables architects and engineers to expand the limits of design, producing distinctive, tailored buildings that were previously prohibitively expensive or challenging to construct using conventional techniques. As a specific example, Rimmer [117] mentioned that maker cities projects have been initiated worldwide, such as in the US, Spain, and the UAE. This 3D printing policy has been utopianly established, especially in Dubai, where construction sector products, consumer products, and medical products were aimed to be made by 3D printing technology.

In the creation of smart cities, 3D printing can expedite the building of infrastructure using IoTs and wireless systems, including on-demand houses [118,119], on-demand bridges [120], and on-demand urban furnishings [121], as shown in Figure 7. The adaptability of 3D printing facilitates on-demand manufacturing, obviating the necessity for extensive warehousing of pre-manufactured components and enhancing supply chain efficacy. Projects such as 3D printed dwellings have shown potential in delivering inexpensive housing alternatives in metropolitan environments for developing countries like Brazil [122]. Although technology readiness levels are 7–8 for an unreinforced 3D printing house where the development of research gap still has to be conducted [122], advancements in robotics and artificial intelligence can fully automate 3D printing systems to create complete structures autonomously, significantly leading to decreased construction time and labor expenses to achieve city livability [123]. Additionally, it should be noted that recycled concrete and nanomaterials can be added to the 3D printing material to improve the sustainability of the holistic construction process [124].

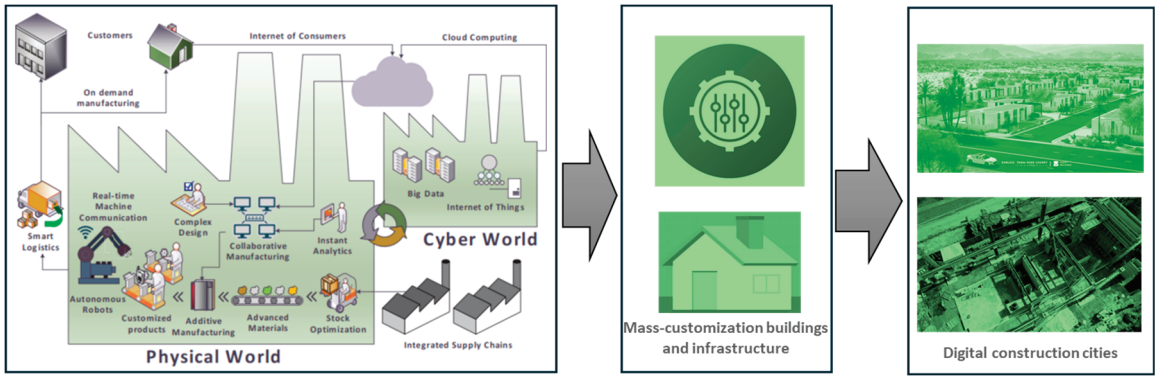


Figure 7. Outline of 3D printing in smart city in Industry 4.0, where mass customization is the key for the future of digital construction cities (source: by author).

Numerous metropolitan areas experience traffic challenges, including congestion, pollution, scheduling difficulties, infrastructure deterioration, and cost reduction concerns for public transportation [125–129]. The swift advancement and deployment of novel IoT technologies have rendered vehicle–infrastructure–pedestrian communication ubiquitous. Technologies such as vehicle to vehicle (V2V), vehicle to infrastructure (V2I), vehicle to pedestrian (V2P), and pedestrian to infrastructure (P2I) have facilitated the development of intelligent transportation systems. Given the prevalence of GPSs and sensor devices in vehicles and the ubiquity of smartphones among drivers, several methodologies utilize GPS data to monitor driver behavior and analyze traffic trends [130,131]. These real-time data are utilized for route planning in programs like Waze and Google Maps, as well as for trip scheduling in public transportation. However, in the context of 3D printing, they can be used for the fast mass customization of manufacturing parts and on-demand attachment of GPSs/sensor devices for systems. Demir et al. [132] advocated that the combined manufacturing and logistics processes of medical 3D printing products and services can be effectively completed during the daytime (8 a.m.–6 p.m.), mitigating the logistic challenge in urban areas.

Nonetheless, the present progress continues to face pragmatic hurdles since the technology is still in its infancy [106]. It is believed that a primary challenge is ensuring effective communication and data exchange across many technologies and platforms. Three-dimensional printing, DTs, and IoT devices must connect seamlessly to provide accurate, real-time insights into building performance. The integration of Industry 4.0 technologies into existing processes may be expensive, especially for smaller construction firms due to smaller budgets and conservative R&D [133,134]. Additionally, as construction and smart city initiatives grow more digitized, they are increasingly vulnerable to cyberattacks [123]. Ensuring the security of IoT devices, data networks, and key infrastructure is imperative to prevent interruptions. The execution of Industry 4.0 requires a workforce skilled in adopting digital technologies, such as AI, robotics, and data analytics [135,136]. The industry must invest in training and upskilling staff to effectively maximize the benefits of these technologies. Finally, governments and industry leaders must devise strategies to alleviate these challenges, potentially through subsidies, incentives, or cooperative platforms [106,137]. Political challenges can be a key parameter for such successful 3D printing technology adopted in the Industry 4.0 context. Indeed, 3D printing can have the main roles in smart manufacturing for smart cities in the context of Industry 4.0.

3.4. Smart 3D Printing Materials for Smart Cities

Smart materials are objects that perceive environmental circumstances and respond accordingly by processing the acquired information [138–140]. The awareness of the environment allows smart materials to be appropriately responsive and reactive. Conversely, non-smart advanced composites often consist of multistructural solid parts that offer multifunctionality. Smart materials can be classified into two categories [141]: the first comprises

materials that exhibit a sensing effect in response to external stimuli, referred to as sensing materials, which can be utilized to create various sensors for detecting environmental changes or gathering information. Sensing materials are now utilized as various sensors, including acoustic emission materials, resistance strain materials, photosensitive materials, moisture-sensitive materials, heat-sensitive materials, and chemical-sensitive materials. The other category comprises materials that react to alterations in external environmental conditions or internal states, sometimes referred to as actuating materials [142–144].

These smart materials often implemented in 3D printing are mainly adopted in filament or ink components, and they are utilized for smart cities and infrastructure as shape memory alloys (SMAs) [145–149], photovoltaic materials [150–155], self-healing materials [156–160], thermochromic materials [161–166], and piezoelectric materials [87,167–174].

The discussion of such 3D printed smart materials is provided herein. For example, Shi et al. [33] illustrated that shell-based ferroelectric metamaterials maintained a high piezoelectric constant at low densities, resulting in enhanced sensitivity to mechanical forces and temperature variations. These materials can be crucial in the advancement of smart city infrastructure, especially in energy-efficient building systems, where immediate reactions to environmental alterations are vital. They also presented an innovative methodology for the design and manufacture of 3D printed ferroelectric metamaterials. These engineered shell-based structures, encompassing spinodoids and diamond shellulars, have programmable piezoelectric and pyroelectric characteristics, rendering them suitable for applications in energy storage, pressure sensing, and thermal systems. This research integrated ferroelectric materials with the 3D printing process, therefore creating new opportunities to improve material efficiency in smart infrastructures, where multifunctionality and customization are essential.

Tuloup et al. [174] elucidated their research on the application of 3D printed lead zirconate titanate (PZT) transducers in SHM systems, which were produced by sophisticated 3D printing methods. The materials are integral to a broader network intended for the real-time detection of structural problems. This method is particularly advantageous in extensive constructions, such as bridges and skyscrapers, where ongoing surveillance is essential to maintain structural integrity and avert disasters. The authors emphasized the advantages of employing printed PZT transducers in comparison to traditional sensors, indicating that they contributed no weight to the construction, were simple to manufacture, and yielded very precise data. The printed PZT transducer can be made from a digital file and subsequently printed and fabricated onsite, where accessibility can be limited. With this customized production and fabrication, the transducer can perform more accurately, even on miniaturized scales. The growing complexity of contemporary infrastructure necessitates regular and non-intrusive SHM to uphold safety and performance criteria in smart cities.

Three-dimensionally printed hydrogels are moisture-responsive materials that may expand or shrink in response to humidity levels for self-healing purposes. These materials exhibit intelligence by adapting or responding to their environmental conditions, enhancing the functioning of passive buildings and rendering them more dynamic and responsive [175,176]. Three-dimensionally printed urethane diacrylate and linear semicrystalline polymers have been formulated for a self-healing polymer ink capable of stretching up to 600% [177].

Smart materials are gaining traction in the 3D printing construction industry, where their unique properties can be applied to improve building performance, improve electrical performance, reduce energy consumption, and enhance the longevity and functionality of infrastructure. Many additives for smart cities can be added to intelligent concretes and asphalts, such as graphite/graphene oxides [178–181], graphene nanomaterials [182–186], and graphene quantum dots [187–190]. The electrical conductivity of the resulting concrete notably increased up to 100% when graphene additives were added [187], with the potential to be adopted in energy autarkic buildings, self-charging pavements for electric vehicles, and energy storage foundations for green wind turbines and tidal power stations.

One of the key applications of 3D printed smart materials in construction is their potential to improve energy efficiency in buildings. Three-dimensionally printed thermochromic windows can change their transparency based on temperature or sunlight intensity [162,191]. When exposed to bright sunlight, these windows become opaque, reducing the amount of heat entering a building and thus lowering cooling costs [192]. On cloudy days or during winter, they remain transparent, maximizing natural light and reducing the need for artificial lighting. Moreover, 3D printed photovoltaic building materials, such as solar panels or solar tiles, are already being used to convert sunlight and heat islands in urban areas into electricity [193–195]. However, new developments in smart photovoltaic materials promise even greater integration with building facades, roofs, and other surfaces, turning buildings into mini power plants that produce clean energy. These materials contribute to the development of zero-energy buildings, a critical goal in smart city planning, where the need for sustainable, energy-efficient infrastructure is paramount.

While 3D printed smart materials hold significant promise, there are several challenges that need to be addressed for widespread adoption, such as cost, scalability, durability, lifespan, standardization, and integration.

- In terms of the economic aspect, smart materials can be costly to produce and may be difficult to scale for large construction projects [196,197]. Hence, a scaling up of production should be performed and find an ameliorative solution.
- Regarding durability performance, some smart materials, such as hydrogels, may degrade over time or under certain environmental conditions, limiting their self-healing effectiveness in long-term applications [198,199]. Fillers or coatings can mitigate the degradation issue.
- Regarding the policy setup, integrating smart materials into existing construction processes requires new proper standards and protocols to ensure compatibility with other materials and systems [200]. This requires collaboration between inventors, city planners, and policymakers to deliberate on the progressive development of future smart cities.

Despite these challenges, the future of smart materials in construction and smart cities is ameliorative. As material science advances and costs decrease, we can expect to see these 3D printed smart materials play an increasingly important role in creating adaptable, sustainable, and intelligent buildings and infrastructure.

3.5. Three-Dimensional Printing in Soft Robotics for Smart Cities

Robots frequently face challenges while functioning in unstructured and densely populated situations. A diverse array of soft robots resembling plant and animal species demonstrate intricate locomotion utilizing pliable frameworks without hard elements [201]. A soft robot relies on the pliability of its subsystems, including the power generator, actuator, sensor, controller, and body including artificial muscles and tissues [202–207].

Soft robotics possesses several applications in smart cities, particularly in infrastructure maintenance, environmental monitoring, and human–robot interaction. Their devices can be utilized for the examination and repair of infrastructure. Conventional inspection techniques frequently utilize inflexible robots or human personnel, which can be unwieldy and pose risks in restricted or intricate settings [208]. Conversely, soft robots can maneuver through confined areas and conform to uneven surfaces, facilitating more effective and safer examinations of bridges, tunnels, and pipelines. Cities can employ 3D printing to create bespoke soft robots that are lightweight and specifically intended for inspection jobs, therefore improving the upkeep of urban infrastructure. Their 3D printed, pliable shapes provide seamless integration into urban settings, reducing interruption while gathering essential data. For instance, 3D printed soft robotic systems may be engineered to emulate plant life, including sensors for environmental surveillance while enhancing urban aesthetics and attracting visitors. The plantoid robot shown in Figure 8 possesses distributed sensors, actuators, and intelligence. Roots are the plant organs that obtain nutrition and provide anchorage. Roots, endowed with multiple sensors at their tips, efficiently investi-

gate their surroundings while continuously adapting, avoiding barriers, penetrating soil, and selectively extracting vital nutrients and water for plant life. Roots mitigate friction and elevated pressures in soil by proliferating new cells at the apical area of the root tips and subsequently absorbing water from the surrounding environment. This capacity increases data collection and develops a more harmonious interaction between technology and nature in urban environments. The model can be applied in several applications in smart cities, including smart farming, in-house gardening, and modern/smart classrooms.

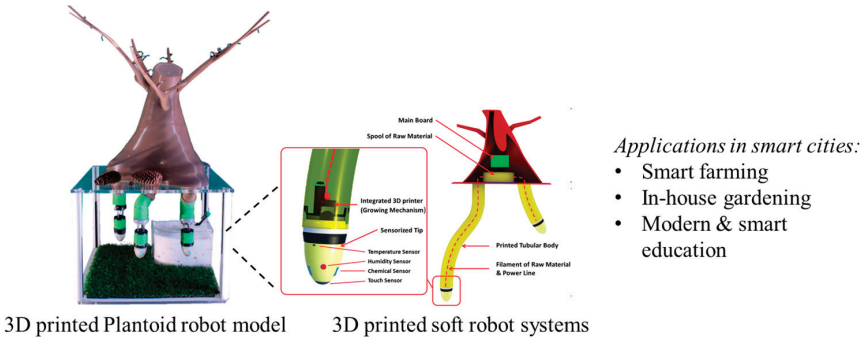


Figure 8. The 3D printed plantoid robot and its systems (source: by author).

Furthermore, soft robotics can enhance human–robot interaction in public environments. Robots intended for social interaction—such as those offering information or help in public spaces—can advantageously utilize soft materials to enhance their approachability and safety for users [209,210]. Another instance is the soft robotic grasper, which is advantageous because of its lightweight nature compared to conventional stiff graspers, and it may streamline the control method for gripping mechanics. A drone capable of flight and perching was constructed, inspired by the resting behaviors of birds and bats in nature, utilizing 3D printed landing gear [211]. Cities may utilize 3D printing technology to provide visually appealing and functionally efficient soft robotic systems, therefore improving the user experience in public areas.

The incorporation of 3D printing in the advancement of soft robotics has several benefits. A primary advantage is the capacity to fabricate intricate, tailored geometries that conventional manufacturing techniques may find challenging to produce. Soft robots sometimes need complex designs to optimize flexibility and functionality. Three-dimensional printing enables designers to swiftly develop and refine ideas, resulting in more inventive and efficient robotic systems customized for specific jobs in urban settings. Furthermore, 3D printing facilitates the utilization of sophisticated materials, such as elastomers and composites, which can improve the functionality of soft robots. These materials may be engineered to offer different levels of flexibility and rigidity, allowing robots to adjust their behavior based on the specific tasks required. A soft robot designed to grasp fragile things can be constructed with a pliable, compliant exterior, yet the same robot can become rigid to execute structural functions as necessary. This adaptability is especially advantageous in smart cities, where the requirements for robotic systems might fluctuate considerably depending on the setting. The cost-effectiveness of 3D printing in soft robotics is a significant benefit. Conventional manufacturing techniques can include substantial setup expenses and extended production lead times. Conversely, 3D printing might diminish production expenses and duration, enabling communities to swiftly implement and modify robotic systems as requirements change. This adaptability is crucial in the dynamic contexts of smart cities, where fast technological progress and evolving urban dynamics may need prompt changes.

Notwithstanding its myriad benefits, the use of 3D printing in soft robotics for smart cities encounters certain hurdles. A primary problem is the longevity and dependability of soft robotic systems [212,213]. Although flexible materials provide distinctive capabilities, they may also jeopardize durability, especially in dense metropolitan regions. To address these problems, continuous research and development programs are necessary to enhance

material characteristics and robotic designs, guaranteeing that soft robots can endure the challenges of urban environments. A further problem is the necessity for multidisciplinary collaboration among engineers, urban planners, and policymakers [214–216]. The success of the plantoid robots (in Figure 8) necessitates complex coordination among customers, agricultural specialists, mechatronics engineers, programmers, and product designers. The aforementioned stakeholders play a crucial role in each component. Furthermore, for the advancement of successful urban smart farming, urban planners and landscape designers must consider land utilization and propose the most effective and urban-friendly configurations for future agricultural, recreational, and tourism locations. Creating efficient soft robotic solutions for smart cities requires an in-depth comprehension of urban dynamics, infrastructural requirements, and community anticipations. Collaborative initiatives may promote innovation and guarantee that soft robotic systems are developed with the end-users' needs in consideration, hence improving their acceptance and efficacy in urban environments. Finally, ethical and social implications accompany the implementation of robotic devices in public areas [217,218]. It is essential that soft robots are created and executed to uphold privacy and improve public safety to ensure their acceptability and success. Involving communities and stakeholders during the design and implementation phases helps mitigate these issues, promoting confidence and collaboration between technology suppliers and urban inhabitants. Ultimately, the breakthrough of soft robotic 3D printing represents a sophisticated vital technology for cutting-edge biological applications in our present and future smart cities.

3.6. Wire-Arc Additive Manufacturing (WAAM) for Smart Cities

WAAM is an innovative 3D printing process that uses an electric arc as a heat source to melt wire material, facilitating the layer-by-layer fabrication of metallic components such as steel, aluminum, alloys, and titanium [219]. Relative to other metal additive manufacturing techniques, due to 3D freeform fabrication, WAAM has reduced cooling rates and increased heat input [220]. This is advantageous for the majority of commercially accessible materials. Moreover, the deposition rates and part dimensions can be enhanced due to the decreased complexity and surface quality of the component when compared to powder-bed-based 3D printing methods.

This novel methodology, as shown in Figure 9, is gaining prominence in the realm of smart cities, where there is an urgent demand for sustainable building methods and sophisticated manufacturing procedures.

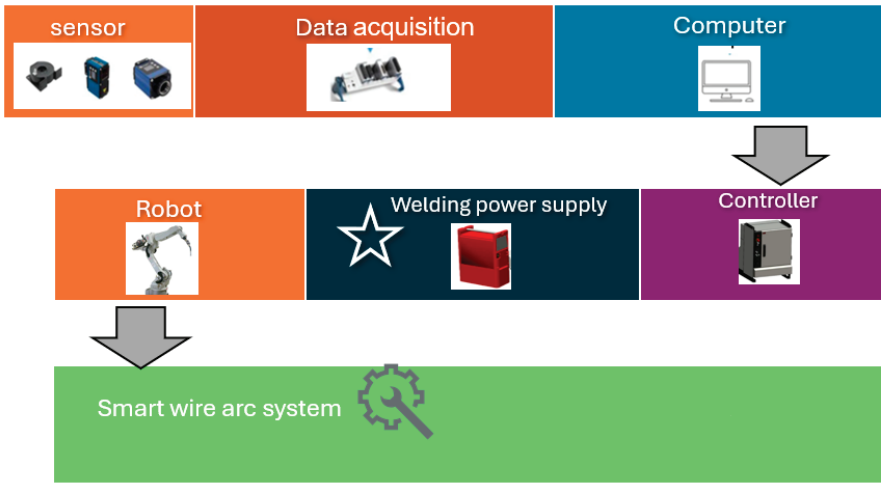


Figure 9. Generic smart wire-arc systems using 3D printing process (source: by author).

One of the primary advantages of WAAM is its capacity to effectively manufacture large-scale components. In contrast to conventional subtractive manufacturing methods that often result in significant waste and prolonged production times, WAAM produces

minimum material waste, hence adhering to the concepts of sustainability and circular economy essential for smart city programs [221–223]. The capacity to fabricate intricate geometries with WAAM can result in creative freeform designs that improve structural performance [224,225]. The layer-by-layer process facilitates the creation of complex forms that are frequently challenging or unattainable by conventional production techniques. This design flexibility is especially advantageous in smart cities, where architects and engineers are prioritizing the optimization of structures for sustainability and efficiency. Furthermore, WAAM can enable the incorporation of multifunctional components, integrating characteristics like thermal insulation or structural reinforcements directly into the produced pieces for smart and greener buildings [226]. The quick prototyping capabilities of WAAM significantly augment its attractiveness for smart city applications. The capacity to rapidly generate and evaluate components facilitates a nimbler design process, enabling engineers to improve designs based on empirical performance data. This iterative methodology is crucial in the dynamic context of smart cities, where infrastructure demands may swiftly evolve owing to variables such as population expansion, technology progress, and climate change [224,227]. WAAM facilitates expedited prototyping, considerably diminishing the time necessary to introduce new designs to the market, thereby improving the overall agility of urban infrastructure development. Notwithstanding its myriad advantages, WAAM has problems that must be resolved for effective adoption in smart cities. A major challenge is the quality monitoring of printed components [228]. Although WAAM can rapidly manufacture large-scale components, it is essential to ensure the mechanical characteristics and structural integrity of these parts. Inconsistencies in the printing process, including fluctuations in heat application and cooling rates, may result in faults that undermine the efficacy of the final product [229]. To alleviate these concerns, continuous research is essential to provide standardized protocols for process monitoring and quality assurance in WAAM systems. A further problem is the necessity for proficient workers to run WAAM systems efficiently [230]. The technique necessitates a profound comprehension of welding and additive manufacturing concepts, which may not be readily accessible within the existing workforce. As smart cities develop, there will be an increasing need for training programs that prepare professionals to run and maintain WAAM systems. In addition, partnerships among educational institutions, businesses, and governments may cultivate a workforce skilled in modern manufacturing processes, promoting innovation and economic development in metropolitan regions [117,231–233]. Many structural applications such as the incorporation of WAAM into current urban infrastructure presents logistical difficulties. Urban settings frequently encounter spatial limitations, necessitating the investigation of modular building techniques to enable the implementation of WAAM technology in confined areas.

WAAM offers a significant possibility for the advancement of smart city development via efficient, sustainable, and creative building methodologies. WAAM can significantly contribute to addressing the changing requirements of urban infrastructure by utilizing its abilities in large-scale component production, fast prototyping, and design adaptability. Addressing issues associated with quality control, workforce development, and logistical integration will be essential for optimizing the technology's potential. As urban areas progress and encounter intricate obstacles, the adoption of WAAM and analogous technologies will be crucial for developing resilient, sustainable, and adaptable urban ecosystems.

3.7. Machine Learning (ML) in 3D Printing for Smart Cities

ML demonstrates the experiential "learning" characteristic of human intelligence, while also possessing the ability to enhance its analysis using computing methods, and it is employed to instruct machines on managing data more effectively [234,235]. Occasionally, after analyzing the data, we are unable to derive the extracted information. In that instance, ML is implemented. The proliferation of datasets has led to an increasing demand for ML. Numerous businesses utilize it to retrieve pertinent data with the key objective of deriving insights from existing data. ML employs many techniques to address data-related

issues. Data scientists emphasize that there is no universal method that is optimal for every situation. ML algorithms for smart cities can be effectively employed due to the large datasets and their selection is contingent upon the specific problem to be addressed, the quantity of variables involved, and the most appropriate model to utilize, among other factors.

The convergence of ML and 3D printing technology is poised to revolutionize various sectors, including construction and urban infrastructure, particularly in the context of smart cities. As urban environments become increasingly complex, leveraging ML in 3D printing can significantly enhance efficiency, sustainability, and adaptability in urban development. One of the primary applications of ML in 3D printing is in the optimization of design and manufacturing processes [74,236,237]. ML algorithms can examine extensive datasets produced during the design process, deriving insights from prior designs and production results to recommend ideal configurations. This data-centric methodology can enhance geometry, material utilization, and structural efficacy. By examining the structural integrity of various designs, ML can forecast which configurations would excel under diverse loads and climatic conditions [238,239]. This feature is especially significant in smart cities, where adaptable infrastructure is essential for addressing evolving urban dynamics. Furthermore, ML algorithms like XGBoost can effectively improve the quality control procedures in the 3D printing process [240,241]. Conventional quality assurance techniques frequently depend on manual inspections and established criteria, which can be labor-intensive, time-consuming, and susceptible to human error [242–244]. Manufacturers may utilize ML algorithms to monitor printing operations in real time, identifying and predicting abnormalities and faults as they arise. This proactive strategy allows prompt modifications to the printing settings, guaranteeing that the final 3D printed products conform to the specified requirements.

Another critical area where ML can contribute to 3D printing in smart cities is predictive maintenance [245,246]. As 3D printing technology becomes more integrated into urban infrastructure, the necessity for continuous maintenance and repair of printed components is paramount. ML algorithms can evaluate past performance data from diverse components to forecast maintenance requirements based on usage patterns and environmental variables. This predictive capacity prolongs the durability of printed buildings, diminishes downtime and maintenance expenses, and eventually fortifies the resilience of urban infrastructure. Moreover, the amalgamation of ML with 3D printing might enable the creation of intelligent, sustainable, and recycled materials that react adaptively to environmental variations [247–250]. For instance, ML algorithms can be employed to engineer materials that alter their characteristics in response to external stimuli, such as temperature, humidity, or mechanical stress. This breakthrough facilitates the development of adaptable structures in smart cities, such as buildings that can optimize temperature and energy consumption or bridges capable of self-repairing small amounts of damage. Since the self-repairing or self-heating system is dependent on time, ML can gather the conditional data from the sites, train, and accurately predict the future variables like the crack width or thickness covering, the strengths, bonding, and surface condition of the repair areas, as well as the timing of the subsequent maintenance. These characteristics correspond with the objectives of sustainability and resilience in urban development.

The integration of ML in 3D printing for smart cities presents several obstacles. A primary problem is the necessity for high-quality, sufficient datasets to train ML models efficiently [251–255]. In several instances, the data necessary for the development of strong algorithms may be scarce or challenging to acquire. To surmount this obstacle, collaboration among academics, businesses, and governments can facilitate the creation of extensive datasets that accurately represent real-world conditions in metropolitan contexts. Moreover, uniformity is required in the amalgamation of ML and 3D printing technologies. As both domains progress swiftly, the formulation of best practices and standards will be essential for guaranteeing interoperability and dependability in smart city applications. Involving stakeholders from many sectors—such as urban planners, engineers, and policymakers—

will be crucial for establishing a unified framework to incorporate new technologies into urban infrastructure. Indeed, ML concepts to analyze the large data of 3D printing processes in smart cities can be applied and integrated well together.

3.8. Three-Dimensional Printing in Structural Health Monitoring (SHM) for Smart Cities

SHM provides valuable insights into a system's behavior by examining its reactions and assessing its present mechanical condition [256,257]. A system may comprise a skyscraper, a bridge, an infrastructure network, or a basic beam. Initially, SHM was utilized for damage detection in airplanes within the field of aerospace engineering and industry. In the past two decades, it has significantly evolved due to sophisticated sensor technologies, substantial improvements in computing, and the integration of structural control methodologies. A conventional simple harmonic motion system has three principal components: (1) sensors, either touch or non-contact, (2) a processing unit (comprising data gathering, transmission, and storage), and (3) a data interpretation system (consisting of diagnostic procedures and information management) [258]. Its purpose in structures is to not just convey its structural integrity, initiating automatic protocols or, if required, requesting human aid, but also to anticipate potential future occurrences [259,260].

SHM denotes the assessment of the state of structures—such as bridges, buildings, and roads—utilizing diverse sensing technologies to guarantee their safety and integrity [261–263]. Conventional SHM frequently depends on integrated sensors and routine evaluations, which are typically labor-intensive and expensive. The use of 3D printing in SHM methods provides novel options to tackle these difficulties. Through the application of 3D printing, municipalities may create bespoke sensors and monitoring apparatus that are lightweight, economical, and simple to install for smart city purposes [264].

A notable benefit of 3D printing in SHM is the capacity to fabricate intricate geometries and structures customized for particular monitoring requirements. Conventional manufacturing techniques may constrain the design of sensor enclosures or supporting structures, but 3D printing facilitates the production of more complex designs [265]. Three-dimensionally printed sensor systems can be included in printed structural components, facilitating smooth integration and reducing interruption to the overall structure. This capability enhances the effectiveness of monitoring systems by ensuring the strategic placement of sensors to gather crucial data. The collected data from 3D printed sensors can be loaded into monitoring or environmental datasets [266–268]. For instance, Kim et al. [267] developed 3D printed vorticella-kirigami-inspired sensors and revealed that these 3D printed wireless sensors can perform 3 times better in terms of out-of-plane strain compared to conventional spring structures with only around 25% strain. In terms of signal detection, the sensors had a sensitivity of 0.80 pF/ms^{-2} through the message queuing telemetry transport protocol, which can be adopted in wireless smartphones and personal computers. This adaptability is especially significant in the realm of smart cities, where the fluidity of urban settings demands prompt reactions to evolving IoT circumstances.

The amalgamation of 3D printing with data analytics and AI significantly augments the efficacy of SHM systems in smart cities. Municipalities may obtain significant insights into structural performance and possible failure risks, like seismic, tornado, and tsunami failure, by integrating real-time data from 3D printed sensors with sophisticated analysis tools [269–272]. Li et al. [273] developed a 3D printed tilt sensor with electrical time domain reflectometry used for SHM in railway level crossings and found promising results with a conservative estimate of sensors' accuracy of $\pm 0.16^{\circ}$, better than the conventional sensor systems. AI systems can analyze extensive datasets, detecting patterns and abnormalities that may signify underlying problems. This data-centric methodology facilitates anticipatory maintenance plans, enabling municipalities to tackle possible issues prior to their escalation into substantial safety risks. In summary, 3D printing is appropriate for the fabrication of SHM sensors and structures, as its mass customization capabilities enable the manufacturing of objects with efficiency and reduced prices.

3.9. Repair Strategies in 3D Printing for Smart Cities

Current infrastructure and buildings have been in service for decades. Many projects have been designed regarding their serviceability design limits like structural and durability performance. This service life of concrete structures, for instance, is stated in "EN 206: Concrete—Requirements, Properties, Production and Compliance" [274] as the duration during which the concrete's condition in the structure meets the operational specifications. The necessity for maintenance seemingly contradicts the belief in the durability of concrete. Numerous manuals on concrete durability and the construction of resilient buildings typically conclude with a chapter titled "How to make repairs" [275]. The seeming contradiction is further demonstrated by the fact that repairs, contingent upon conditions, aim to preserve or restore the durability of a building's construction [276,277]. Hence, repair strategies are key activities of restoration of buildings and infrastructure. This can be integrated with new innovation for smart cities.

As urban areas confront issues associated with deteriorating infrastructure, resource depletion, and sustainability, 3D printing provides novel repair methodologies that improve resilience and efficiency. Three-dimensional printing enables a prominent repair strategy that involves the on-demand production of replacement components [278,279]. Conventional supply chains sometimes entail extended lead times and significant logistical difficulties, especially for specialty components. Conversely, 3D printing allows towns to produce components locally, therefore diminishing transportation emissions and expenses [280]. This localized production strategy accelerates maintenance operations and enables towns to retain a supply of essential components customized to their individual requirements. Urban regions can utilize 3D printing to manufacture replacement components for public infrastructure, such as lamps, benches, or signage, therefore reducing service disruptions and improving the overall efficacy of public places [281]. Engineers may develop components that are more resilient and better equipped to endure environmental difficulties by leveraging modern and ecologically adaptive materials and implementing design optimization methods. Alongside on-demand manufacture, adaptable design methodologies are essential to the efficacy of 3D printed repairs. In this on-demand repair, new 3D printing materials with faster hardening are required [282,283]. As examples of cement-based materials, alternative binders and chemical additives like calcium aluminate, calcium sulfoaluminate, and accelerator were introduced for such applications [284–287]. Kumar [288] also used polyvinylidene fluoride (PVDF)–graphene–manganese (Mn)-doped zinc oxide (ZnO) and PVDF-based 3D printable solutions for smart maintenance of non-structural cracks of heritage structures.

Predictive maintenance enabled by the IoT enhances the efficacy of repair solutions in 3D printing [289–291]. IoT sensors may be integrated into infrastructure to perpetually assess performance and identify possible faults prior to their escalation. This proactive strategy allows communities to plan repairs effectively, minimizing the chances of unforeseen failures and enhancing resource distribution. For instance, intelligent streetlight systems with sensors may notify maintenance personnel when components need repair, facilitating prompt and precise interventions. The collaboration of 3D printing and IoT technology highlights the potential for a data-centric approach to urban infrastructure management [292,293]. Furthermore, promoting collaboration among stakeholders is essential for the development and execution of successful repair plans in 3D printing. Involving city planners, engineers, architects, and the local community cultivates an atmosphere conducive to the flourishing of ideas and inventions. Sustainability constitutes a crucial element of repair solutions in 3D printing for smart cities [10,294]. The circular economy paradigm, which prioritizes reuse and recycling, may be included in the design and repair processes. For example, wasted plastic trash may be transformed into filament for FDM 3D printing, diminishing landfill contributions by >90% while offering a useful resource for infrastructure restoration [295,296]. Another example is adding polymer and additive waste [297–299] or using low-energy intensive binders [230–302] and reinforcement [72,303–308] in 3D printed concrete/polymer matrix in concrete structures for increasing flexibility and promoting a circular economy. This method not only reduces

trash but also fosters a culture of sustainability in urban environments. WAAM is an innovative 3D printing process that uses an electric arc as a heat source to melt wire material, facilitating the layer-by-layer fabrication of metallic components such as steel, aluminum, and titanium. This novel methodology, as shown in Figure 8, is gaining prominence in the realm of smart cities, where there is an urgent demand for sustainable building methods and sophisticated manufacturing procedures. Without doubt, 3D printing repair strategies with freeform printing and adaptive fabrication serve well for restoring and maintaining buildings and infrastructure for smarter cities.

4. Conclusions

The article conducted qualitative analysis and a systematic review of the contribution of 3D printing to smart cities, which underscores the substantial contributions of the study to academic debate and practical understanding from state-of-the-art publications.

- The review elucidates the interaction of DTs, smart materials, Industry 4.0, soft robotics, ML, SHM, WAAM, and repair strategies for construction, urban planning, smart infrastructure, and environmental sustainability through the integration of many views. All of the mentioned areas have research gaps and roadmaps to notably improve the future of smart cities.
- DTs and 3D printing can be utilized in producing building elements in existing buildings.
- Industry 4.0 presents the future of mass-customized digital construction.
- Smart materials in 3D printing offer distinct benefits like improved building, electrical, and functional performance, reduced energy consumption, and enhanced durability.
- ML can evaluate past performance data from diverse components from existing smart buildings to forecast maintenance requirements as well as predict and control environmental conditions and cities for more livable communities.
- Soft robotics are key for human interaction with automated machines, with implementation varying from medicine to gardening and farming.
- SHM is the key to inspecting and maintain safety in cities with less cost and more security regarding unpredictable, diverse events.
- Repair strategies of existing building structures can be made easier and more effective by the WAAM approach, yielding strategized planning of repairs for communities, minimizing the possibility of failure, and enhancing resource distribution.

The analysis delineates synergies, obstacles, constraints, and research gaps while providing prospective recommendations to enhance future investigations. Academics, practitioners, and policymakers can follow the recommendations and insights of this work to implement powerful technology for local communities. These insights facilitate the development of more resilient, technologically sophisticated, and ecologically sustainable urban landscapes. Despite the promising advancements, ubiquitous challenges remain in terms of cost, scalability, established policies and regulations, and the need for interdisciplinary collaboration among engineers, urban planners, and policymakers. This paper offers a detailed framework for academics, practitioners, and policymakers to inform the development of more intelligent, sustainable urban environments. These insights have profound implications for researchers, practitioners, and policymakers, providing a roadmap for fostering resiliently designed, technologically advanced, and environmentally conscious urban environments.

Funding: This project was also funded by National Research Council of Thailand (NRCT) and Chulalongkorn University (N42A660629) and by Thailand Science Research and Innovation Fund Chulalongkorn University.

Data Availability Statement: Data will be made available on request.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Dixon, T.; Eames, M. Sustainable urban development to 2050: Complex transitions in the built environment of cities. In *Urban Retrofitting for Sustainability*; Routledge: London, UK, 2014; pp. 19–47.
2. Asif, M.; Naeem, G.; Khalid, M. Digitalization for sustainable buildings: Technologies, applications, potential, and challenges. *J. Clean. Prod.* **2024**, *450*, 141814. [CrossRef]
3. Žujović, M.; Obradović, R.; Rakonjac, I.; Milošević, J. 3D printing technologies in architectural design and construction: A systematic literature review. *Buildings* **2022**, *12*, 1319. [CrossRef]
4. Darwish, L.R.; El-Wakad, M.T.; Farag, M.M. Towards sustainable industry 4.0: A green real-time IIoT multitask scheduling architecture for distributed 3D printing services. *J. Manuf. Syst.* **2021**, *61*, 196–209. [CrossRef]
5. Li, L.; Lange, K.W. Planning principles for integrating community empowerment into zero-net carbon transformation. *Smart Cities* **2022**, *6*, 100–122. [CrossRef]
6. Loy, J.; Novak, J.I. Additive manufacturing for a dematerialized economy. In *Sustainable Manufacturing and Design*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 19–45.
7. Wang, Q.-C.; Yu, S.-N.; Chen, Z.-X.; Weng, Y.-W.; Xue, J.; Liu, X. Promoting additive construction in fast-developing areas: An analysis of policies and stakeholder perspectives. *Dev. Built Environ.* **2023**, *16*, 100271. [CrossRef]
8. Sachs, J.D.; Lafortune, G.; Fuller, G.; Drumm, E. *Sustainable Development Report 2023: Implementing the SDG Stimulus*; Dublin University Press: Dublin, Ireland, 2023.
9. Bamigboye, G.O.; Bassey, D.E.; Olukanni, D.O.; Ngene, B.U.; Adegoke, D.; Odetoyan, A.O.; Kareem, M.A.; Enabulele, D.O.; Nworgu, A.T. Waste materials in highway applications: An overview on generation and utilization implications on sustainability. *J. Clean. Prod.* **2021**, *283*, 124581. [CrossRef]
10. Ahmed, G.H. A review of "3D concrete printing": Materials and process characterization, economic considerations and environmental sustainability. *J. Build. Eng.* **2023**, *66*, 105863. [CrossRef]
11. Muriungi, K.; Chumo, K.P. Building Future-Proof, Sustainable, and Innovative Startups in Africa Powered by Smart Cities. In *Green Computing for Sustainable Smart Cities*; CRC Press: Boca Raton, FL, USA, 2024; pp. 172–205.
12. Kumar, K.; Zindani, D.; Davim, J.P. *Rapid Prototyping, Rapid Tooling and Reverse Engineering: From Biological Models to 3D Bioprinters*; Walter de Gruyter GmbH & Co KG: Berlin, Germany, 2020; Volume 5.
13. Jiramarootapong, P.; Prasittisopin, L.; Snguanyat, C.; Tanapornraweekit, G.; Tangtermsirikul, S. Load carrying capacity and failure mode of 3D printing mortar wall panel under axial compression loading. In *Second RILEM International Conference on Concrete and Digital Fabrication: Digital Concrete*; Springer: Cham, Switzerland, 2020; pp. 646–657.
14. Prasittisopin, L.; Sakdanaraseth, T.; Horayangkura, V. Design and construction method of a 3D concrete printing self-supporting curvilinear pavilion. *J. Archit. Eng.* **2021**, *27*, 05021006. [CrossRef]
15. Sadakorn, W.; Prasertsuk, S.; Prasittisopin, L. 3D Cement Printing: DFMA Guideline of Patterned Load-Bearing Walls for Small Residential Units. In *International Conference on Civil Engineering and Architecture, 2022*; Springer: Singapore, 2022; pp. 19–28.
16. Bhanye, J.; Lehobo, M.T.; Mocwagae, K.; Shayamunda, R. Strategies for sustainable innovative affordable housing (SIAH) for low income families in africa: A rapid review study. *Discov. Sustain.* **2024**, *5*, 157. [CrossRef]
17. Berjozkina, G.; Karami, R. 3D printing in tourism: An answer to sustainability challenges? *Worldw. Hosp. Tour. Themes* **2021**, *13*, 773–788.
18. Ayad, T.H.; Shehata, A.E.S. The Polential Impact of 3D printing technologies on travel and hospitality industry. *J. Assoc. Arab Univ. Tour. Hosp.* **2014**, *11*, 49–58. [CrossRef]
19. Schelly, C.; Anzalone, G.; Wijnen, B.; Pearce, J.M. Open-source 3-D printing technologies for education: Bringing additive manufacturing to the classroom. *J. Vis. Lang. Comput.* **2015**, *28*, 226–237. [CrossRef]
20. Maloy, R.; Kommers, S.; Malinowski, A.; LaRoche, I. 3D modeling and printing in history/social studies classrooms: Initial lessons and insights. *Contemp. Issues Technol. Teach. Educ.* **2017**, *17*, 229–249.
21. Houser, K. *World's Largest 3D-Printed Affordable Housing Project Launches in Kenya*; Freethink: New York, NY, USA, 2023.
22. Faleschini, F.; Trento, D.; Masoomi, M.; Pellegrino, C.; Zanini, M.A. Sustainable mixes for 3D printing of earth-based constructions. *Constr. Build. Mater.* **2023**, *398*, 132496. [CrossRef]
23. Tetiranont, S.; Sadakorn, W.; Rugkhapan, N.T.; Prasittisopin, L. Enhancing sustainable railway station design in tropical climates: Insights from Thailand's architectural theses and case studies. *Buildings* **2024**, *14*, 829. [CrossRef]
24. Mikula, K.; Skrzypczak, D.; Izydorczyk, G.; Warchoł, J.; Moustakas, K.; Chojnacka, K.; Witek-Krowiak, A. 3D printing filament as a second life of waste plastics—A review. *Environ. Sci. Pollut. Res.* **2021**, *28*, 12321–12333. [CrossRef]
25. Madhu, N.R.; Erfani, H.; Jadoun, S.; Amir, M.; Thiagarajan, Y.; Chauhan, N.P.S. Fused deposition modelling approach using 3D printing and recycled industrial materials for a sustainable environment: A review. *Int. J. Adv. Manuf. Technol.* **2022**, *122*, 2125–2138. [CrossRef]
26. Bartolomei, S.S.; da Silva, F.L.F.; de Moura, E.A.B.; Wiebeck, H. Recycling expanded polystyrene with a biodegradable solvent to manufacture 3D printed prototypes and finishing materials for construction. *J. Polym. Environ.* **2022**, *30*, 3701–3717. [CrossRef]
27. Bremer, M.; Janoschek, L.; Kaschta, D.; Schneider, N.; Wahl, M. Influence of plastic recycling—A feasibility study for additive manufacturing using glycol modified polyethylene terephthalate (PETG). *SN Appl. Sci.* **2022**, *4*, 156. [CrossRef]
28. de Rubeis, T. 3D-Printed Blocks: Thermal Performance Analysis and Opportunities for Insulating Materials. *Sustainability* **2022**, *14*, 1077. [CrossRef]

29. Sun, J.; Xiao, J.; Li, Z.; Feng, X. Experimental study on the thermal performance of a 3D printed concrete prototype building. *Energy Build.* **2021**, *241*, 110965. [CrossRef]
30. Sánchez-Corcuera, R.; Nuñez-Marcos, A.; Sesma-Solance, J.; Bilbao-Jayo, A.; Mulero, R.; Zulaika, U.; Azkune, G.; Almeida, A. Smart cities survey: Technologies, application domains and challenges for the cities of the future. *Int. J. Distrib. Sens. Netw.* **2019**, *15*, 1550147719853984. [CrossRef]
31. O'Dea, R.E.; Lagisz, M.; Jennions, M.D.; Koricheva, J.; Noble, D.W.; Parker, T.H.; Gurevitch, J.; Page, M.J.; Stewart, G.; Moher, D. Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: A PRISMA extension. *Biol. Rev.* **2021**, *96*, 1695–1722. [CrossRef] [PubMed]
32. Zastrow, M. The new 3D printing. *Nature* **2020**, *578*, 20–23. [CrossRef] [PubMed]
33. Shi, J.; Ju, K.; Chen, H.; Mirabolghasemi, A.; Akhtar, S.; Sasmito, A.; Akbarzadeh, A. 3D printed architected shell-based ferroelectric metamaterials with programmable piezoelectric and pyroelectric properties. *Nano Energy* **2024**, *123*, 109385. [CrossRef]
34. Mu, H.; He, F.; Yuan, L.; Hatamian, H.; Commins, P.; Pan, Z. Online distortion simulation using generative machine learning models: A step toward digital twin of metallic additive manufacturing. *J. Ind. Inf. Integr.* **2024**, *38*, 100563. [CrossRef]
35. Salaimanimagudam, M.P.; Jayaprakash, J. Selection of digital fabrication technique in the construction industry—A multi-criteria decision-making approach. *Front. Struct. Civ. Eng.* **2024**, *18*, 977–997. [CrossRef]
36. Paunikar, S.; Galanopoulos, G.; Rebillat, M.; Wirth, I.; Monteiro, E.; Margerit, P.; Mechbal, N. Printed PZT Transducers Network for the Structural Health Monitoring of Foreign Object Damage Composite Panel. In Proceedings of the 11th European Workshop on Structural Health Monitoring, EWSHM, Potsdam, Germany, 10–13 June 2024.
37. Rojek, I.; Marciniak, T.; Mikołajewski, D. Digital twins in 3D printing processes using artificial intelligence. *Electronics* **2024**, *13*, 3550. [CrossRef]
38. Mattera, G.; Nele, L.; Paolella, D. Monitoring and control the wire arc additive manufacturing process using artificial intelligence techniques: A review. *J. Intell. Manuf.* **2024**, *35*, 467–497. [CrossRef]
39. Corzo, D.; Alexandre, E.B.; Alshareef, Y.; Bokhari, F.; Xin, Y.; Zhang, Y.; Kosel, J.; Bryant, D.; Lubineau, G.; Baran, D. Cure-on-demand 3D printing of complex geometries for enhanced tactile sensing in soft robotics and extended reality. *Mater. Today* **2024**, *78*, 20–31. [CrossRef]
40. Shufrin, I.; Pasternak, E.; Dyskin, A. Environmentally friendly smart construction—Review of recent developments and opportunities. *Appl. Sci.* **2023**, *13*, 12891. [CrossRef]
41. Botero-Valencia, J.; Martinez-Perez, A.; Hernández-García, R.; Castano-Londono, L. Exploring spatial patterns in sensor data for humidity, temperature, and RSSI measurements. *Data* **2023**, *8*, 82. [CrossRef]
42. Engel, L.; Khouri, D.; Lomakin, K.; Hofmann, A.; Kleinlein, M.; Ullmann, I.; Vossiek, M.; Gold, G. 3D printed hemispherically radiating antenna for broadband millimeter wave applications. *IEEE Open J. Antennas Propag.* **2023**, *4*, 558–570. [CrossRef]
43. Lu, X.; Zhang, G.; Chiumenti, M.; Cervera, M.; Slimani, M.; Ma, L.; Wei, L.; Lin, X. Smart-substrate: A novel structural design to avert residual stress accretion in directed energy deposition additive manufacturing. *Virtual Phys. Prototyp.* **2023**, *18*, e2246041. [CrossRef]
44. Raymond, L.; Bandala, E.; Coulter, R.; Valentin, N.; Mitchell, K.; Hua, W.; Zhang, C.; Zhao, D.; Jin, Y. Hybrid direct ink writing/embedded three-dimensional printing of smart hinge from shape memory polymer. *Manuf. Lett.* **2023**, *35*, 609–619. [CrossRef]
45. Stano, G.; Ovy, S.M.A.I.; Edwards, J.R.; Cianchetti, M.; Percoco, G.; Tadesse, Y. One-shot additive manufacturing of robotic finger with embedded sensing and actuation. *Int. J. Adv. Manuf. Technol.* **2023**, *124*, 467–485. [CrossRef]
46. Keller, S.; Stein, M.; Ilic, O. Material extrusion on an ultrasonic air bed for 3D printing. *J. Vib. Acoust.* **2023**, *145*, 061001. [CrossRef]
47. Mu, H.; He, F.; Yuan, L.; Commins, P.; Wang, H.; Pan, Z. Toward a smart wire arc additive manufacturing system: A review on current developments and a framework of digital twin. *J. Manuf. Syst.* **2023**, *67*, 174–189. [CrossRef]
48. Dhall, S.; Rab, S.; Pal, S.K.; Javaid, M.; Khan, A.A.; Haleem, A. Identifying the feasibility of 'travelator roads' for modern-era sustainable transportation and its prototyping using additive manufacturing. *Sustain. Oper. Comput.* **2023**, *4*, 119–129. [CrossRef]
49. Kumar, V.; Singh, R.; Ahuja, I.S. Smart Thermoplastics for Maintenance and Repair of Heritage Structures. *Encycl. Mater. Plast. Polym.* **2022**, *1*, 524–532.
50. Binder, M.; Stapff, V.; Heinig, A.; Schmitt, M.; Seidel, C.; Reinhart, G. Additive manufacturing of a passive, sensor-monitored 16MnCr5 steel gear incorporating a wireless signal transmission system. *Procedia CIRP* **2022**, *107*, 505–510. [CrossRef]
51. Ranjan, R.; Kumar, D.; Kundu, M.; Chandra Moi, S. A critical review on classification of materials used in 3D printing process. *Mater. Today Proc.* **2022**, *61*, 43–49. [CrossRef]
52. Tapas, N.; Belikovetsky, S.; Longo, F.; Puliafito, A.; Shabtai, A.; Elovici, Y. 3D Marketplace: Distributed Attestation of 3D Designs on Blockchain. In Proceedings of the 2022 IEEE International Conference on Smart Computing, SMARTCOMP, Helsinki, Finland, 20–24 June 2022; pp. 311–316.
53. Alfattni, R. Comprehensive study on materials used in different types of additive manufacturing and their applications. *Int. J. Math. Eng. Manag. Sci.* **2022**, *7*, 92–114. [CrossRef]
54. Safaee, S.; Otero, A.; Fei, M.; Liu, T.; Zhang, J.; Chen, R.K. Particle-resin systems for additive manufacturing of rigid and elastic magnetic polymeric composites. *Addit. Manuf.* **2022**, *51*, 102587. [CrossRef]
55. Jian, B.; Demoly, F.; Zhang, Y.; Qi, H.J.; André, J.C.; Gomes, S. Origami-based design for 4D printing of 3D support-free hollow structures. *Engineering* **2022**, *12*, 70–82. [CrossRef]

56. Gupta, S.; Bag, S.; Modgil, S.; Beatriz Lopes de Sousa Jabbour, A.; Kumar, A. Examining the influence of big data analytics and additive manufacturing on supply chain risk control and resilience: An empirical study. *Comput. Ind. Eng.* **2022**, *172*, 108629. [CrossRef]
57. Liu, C.; Chang, J. Electronics—A First Course for Printed Circuit Board Design. In Proceedings of the ASEE Annual Conference and Exposition, Conference Proceedings, Minneapolis, MN, USA, 26–29 June 2022.
58. Kyvelou, P.; Buchanan, C.; Gardner, L. Numerical simulation and evaluation of the world's first metal additively manufactured bridge. *Structures* **2022**, *42*, 405–416. [CrossRef]
59. Lubin, L.; Nwokedi, I.; Diwan, Y.; Moreno, O.; Kippelen, B. Luminaire for Connected Lighting System with Spectrum that Mimics Natural Light. In Proceedings of the 2022 Opportunity Research Scholars Symposium, ORSS, Atlanta, GA, USA, 27 April 2022; pp. 49–52.
60. Reisch, R.T.; Hauser, T.; Lutz, B.; Tsakpinis, A.; Winter, D.; Kamps, T.; Knoll, A. Context awareness in process monitoring of additive manufacturing using a digital twin. *Int. J. Adv. Manuf. Technol.* **2022**, *119*, 3483–3500. [CrossRef]
61. Rubino, M.; Weng, M.; Chen, J.; Saptarshi, S.; Francisco, M.; Francisco, A.; Zhou, C.; Sun, H.; Xu, W. A Campus Prototype of Interactive Digital Twin in Cyber Manufacturing. In Proceedings of the SenSys 2022: Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems, Boston, MA, USA, 6–9 November 2022; pp. 758–759.
62. Na, S.; Kim, S.; Moon, S. Additive manufacturing (3D Printing)-applied construction: Smart node system for an irregular building façade. *J. Build. Eng.* **2022**, *56*, 104743. [CrossRef]
63. Ferretti, P.; Santi, G.M.; Leon-Cardenas, C.; Freddi, M.; Donnici, G.; Frizziero, L.; Liverani, A. Molds with advanced materials for carbon fiber manufacturing with 3d printing technology. *Polymers* **2021**, *13*, 3700. [CrossRef]
64. Eleftheriadis, G.K.; Genina, N.; Boetker, J.; Rantanen, J. Modular design principle based on compartmental drug delivery systems. *Adv. Drug Deliv. Rev.* **2021**, *178*, 113921. [CrossRef] [PubMed]
65. Cohn, D.; Zarek, M.; Elyashiv, A.; Sbitan, M.A.; Sharma, V.; Ramanujan, R.V. Remotely triggered morphing behavior of additively manufactured thermoset polymer-magnetic nanoparticle composite structures. *Smart Mater. Struct.* **2021**, *30*, 045022. [CrossRef]
66. Alwis, L.S.M.; Bremer, K.; Roth, B. Fiber optic sensors embedded in textile-reinforced concrete for smart structural health monitoring: A review. *Sensors* **2021**, *21*, 4948. [CrossRef] [PubMed]
67. Li, J.; Yang, F.; Long, Y.; Dong, Y.; Wang, Y.; Wang, X. Bulk ferroelectric metamaterial with enhanced piezoelectric and biomimetic mechanical properties from additive manufacturing. *ACS Nano* **2021**, *15*, 14903–14914. [CrossRef]
68. Villacres, J.; Guaman, R.; Menendez, O.; Auat Cheein, F. 3D Printing Deformation Estimation Using Artificial Vision Strategies for Smart-Construction. In Proceedings of the IECON 2021—47th Annual Conference of the IEEE Industrial Electronics Society, Toronto, ON, Canada, 13–16 October 2021.
69. Mileti, I.; Cortese, L.; Del Prete, Z.; Palermo, E. Reproducibility and embedding effects on static performace of 3D printed strain gauges. In Proceedings of the 2021 IEEE International Workshop on Metrology for Industry 4.0 and IoT, MetroInd 4.0 and IoT, Rome, Italy, 7–9 June 2021; pp. 499–504.
70. Zhang, X.; Liou, F. Introduction to additive manufacturing. In *Additive Manufacturing*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 1–31.
71. Surie, G. DIGITAL TECHNOLOGIES AND THE WORLD OF THE FUTURE: IMPLICATIONS FOR THE MANAGEMENT OF TECHNOLOGY. In Proceedings of the 30th International Conference of the International Association for Management of Technology, IAMOT 2021—MOT for the World of the Future, Cairo, Egypt, 19–23 September 2021; pp. 173–189.
72. Vlachakis, C.; Perry, M.; Biondi, L.; McAlorum, J. 3D printed temperature-sensing repairs for concrete structures. *Addit. Manuf.* **2020**, *34*, 101238. [CrossRef]
73. Fan, J.; Gonzalez, D.; Garcia, J.; Newell, B.; Nawrocki, R.A. In The effects of additive manufacturing and electric poling techniques on PVdF thin films: Towards 3D printed functional materials. In Proceedings of the ASME 2020 Conference on Smart Materials, Adaptive Structures and Intelligent Systems, SMASIS, Online, 15 September 2020.
74. Jin, Z.; Zhang, Z.; Demir, K.; Gu, G.X. Machine learning for advanced additive manufacturing. *Matter* **2020**, *3*, 1541–1556. [CrossRef]
75. Hehr, A.; Norfolk, M.; Kominsky, D.; Boulanger, A.; Davis, M.; Boulware, P. Smart build-plate for metal additive manufacturing processes. *Sensors* **2020**, *20*, 360. [CrossRef]
76. Chen, X.; Wang, D.; Jiang, T.; Xiao, H. Realtime Control-oriented Modeling and Disturbance Parameterization for Smart and Reliable Powder Bed Fusion Additive Manufacturing. In Proceedings of the Solid Freeform Fabrication 2018: Proceedings of the 29th Annual International Solid Freeform Fabrication Symposium—An Additive Manufacturing Conference, SFF, Austin, TX, USA, 13–15 August 2018; pp. 2335–2348.
77. Castle, H. Disrupting from the Inside: UK archipreneurs. *Arch. Des.* **2020**, *90*, 40–49. [CrossRef]
78. Chen, K.W.; Tsai, M.J. Multi-Nozzle Pneumatic Extrusion Based Additive Manufacturing System for Fabricating a Sandwich Structure with Soft and Hard Material. In Proceedings of the 2019 International Conference on Machine Learning and Cybernetics (ICMLC), Kobe, Japan, 7–10 July 2019.
79. Zhang, J.; Wang, J.; Dong, S.; Yu, X.; Han, B. A review of the current progress and application of 3D printed concrete. *Compos. Part A Appl. Sci. Manuf.* **2019**, *125*, 105533. [CrossRef]
80. Hamidi, F.; Aslani, F. Additive manufacturing of cementitious composites: Materials, methods, potentials, and challenges. *Constr. Build. Mater.* **2019**, *218*, 582–609. [CrossRef]

81. Yoon, J. SMP Prototype design and fabrication for thermo-responsive façade elements. *J. Facade Des. Eng.* **2019**, *7*, 41–61.
82. Tay, Y.W.D.; Panda, B.N.; Ting, G.H.A.; Ahamed, N.M.N.; Tan, M.J.; Chua, C.K. 3D printing for sustainable construction. In *Industry 4.0—Shaping the Future of the Digital World*; CRC Press: Boca Raton, FL, USA, 2019; pp. 119–123.
83. Yu, Y.; Mikkilineni, A.K.; Killough, S.M.; Kuruganti, T.; Joshi, P.C.; Hu, A. Direct-Write Printed Current Sensor for Load Monitoring Applications. In Proceedings of the 2019 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference, ISGT, Washington, DC, USA, 18–21 February 2019.
84. Chhetri, S.R.; Faezi, S.; Canedo, A.; Faruque, M.A.A. In QUILT: Quality inference from living digital twins in IoT-enabled manufacturing systems. In Proceedings of the 2019 Internet of Things Design and Implementation, Montreal, QC, Canada, 15–18 April 2019; pp. 237–248.
85. Mallineni, S.S.K.; Dong, Y.; Behlow, H.; Rao, A.M.; Podila, R. A wireless triboelectric nanogenerator. *Adv. Energy Mater.* **2018**, *8*, 1702736. [CrossRef]
86. Arrizubieta, J.I.; Ruiz, J.E.; Martinez, S.; Ukar, E.; Lamikiz, A. Intelligent nozzle design for the laser metal deposition process in the industry 4.0. *Procedia Manuf.* **2017**, *13*, 1237–1244. [CrossRef]
87. Emon, M.O.F.; Choi, J.W. In A prelimiary study on 3D printed smart insoles with stretchable piezoresistive sensors for plantar pressure monitoring. In Proceedings of the ASME 2017 International Mechanical Engineering Congress and Exposition, Tampa, FL, USA, 3–9 November 2017.
88. Mathew, J.; Hauser, C.; Stoll, P.; Kenel, C.; Polyzos, D.; Havermann, D.; Macpherson, W.N.; Hand, D.P.; Leinenbach, C.; Spierings, A.; et al. Integrating Fiber Fabry-Perot Cavity Sensor into 3-D Printed Metal Components for Extreme High-Temperature Monitoring Applications. *IEEE Sens. J.* **2017**, *17*, 4107–4114. [CrossRef]
89. Paritala, P.K.; Manchikatla, S.; Yarlagadda, P.K.D.V. Digital manufacturing-applications past, current, and future trends. *Procedia Eng.* **2017**, *174*, 982–991. [CrossRef]
90. Gooding, J.; Mahoney, J.F.; Fields, T.D. 3D printed strain gauge geometry and orientation for embedded sensing. In Proceedings of the 58th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Grapevine, TX, USA, 9–13 January 2017.
91. Jackson, R.; Curran, S.; Chambon, P.; Post, B.; Love, L.; Wagner, R.; Ozpineci, B.; Chinthavali, M.; Starke, M.; Green, J.; et al. Overview of the oak ridge national laboratory advanced manufacturing integrated energy demonstration project: Case study of additive manufacturing as a tool to enable rapid innovation in integrated energy systems. In Proceedings of the ASME 2016 International Mechanical Engineering Congress and Exposition, Phoenix, AZ, USA, 11–17 November 2016.
92. Shafaroudi, A.A.; Rankohi, S. Multifunctional and multiphysics materials as load-bearing structural components. In Proceedings of the 2016 Annual Conference of the Canadian Society for Civil Engineering, London, UK, 1–4 June 2016; pp. 2108–2116.
93. Ghazanfari, A.; Li, W.; Leu, M.C.; Zhuang, Y.; Huang, J. Advanced ceramic components with embedded sapphire optical fiber sensors for high temperature applications. *Mater. Des.* **2016**, *112*, 197–206. [CrossRef]
94. Kimionis, J.; Isakov, M.; Koh, B.S.; Georgiadis, A.; Tentzeris, M.M. 3D-Printed origami packaging with inkjet-printed antennas for RF harvesting sensors. *IEEE Trans. Microw. Theory Tech.* **2015**, *63*, 4521–4532. [CrossRef]
95. Korenhof, P.; Blok, V.; Kloppenburg, S. Steering representations—Towards a critical understanding of digital twins. *Philos. Technol.* **2021**, *34*, 1751–1773. [CrossRef]
96. Grübel, J.; Thrash, T.; Aguilar, L.; Gath-Morad, M.; Chatain, J.; Sumner, R.W.; Hölscher, C.; Schinazi, V.R. The hitchhiker's guide to fused twins: A review of access to digital twins in situ in smart cities. *Remote Sens.* **2022**, *14*, 3095. [CrossRef]
97. Vatani, M.; Lu, Y.; Engeberg, E.D.; Choi, J.W. Combined 3D printing technologies and material for fabrication of tactile sensors. *Int. J. Precis. Eng. Manuf.* **2015**, *16*, 1375–1383. [CrossRef]
98. Tuvayanond, W.; Prasittisopin, L. Design for manufacture and assembly of digital fabrication and additive manufacturing in construction: A review. *Buildings* **2023**, *13*, 429. [CrossRef]
99. Kristombu Baduge, S.; Navaratnam, S.; Abu-Zidan, Y.; McCormack, T.; Nguyen, K.; Mendis, P.; Zhang, G.; Aye, L. Improving performance of additive manufactured (3D printed) concrete: A review on material mix design, processing, interlayer bonding, and reinforcing methods. *Structures* **2021**, *29*, 1597–1609. [CrossRef]
100. Sadakorn, W.; Prasertsuk, S.; Prasittisopin, L. Improving the structural efficiency of textured three-dimensional concrete printing wall by architectural design. *Front. Struct. Civ. Eng.* **2024**, *18*, 699–715. [CrossRef]
101. Bibri, S.E.; Huang, J.; Jagatheesaperumal, S.K.; Krogstie, J. The synergistic interplay of artificial intelligence and digital twin in environmentally planning sustainable smart cities: A comprehensive systematic review. *Environ. Sci. Ecotechnol.* **2024**, *20*, 100433. [CrossRef] [PubMed]
102. Sagl, G.; Resch, B.; Blaschke, T. Contextual sensing: Integrating contextual information with human and technical geo-sensor information for smart cities. *Sensors* **2015**, *15*, 17013–17035. [CrossRef] [PubMed]
103. Sangawongse, S.; Ruangrit, V. Towards a GIS-based urban information system to plan a smarter Chiang Mai. *Nakhara J. Environ. Des. Plan.* **2015**, *11*, 1.
104. Attaran, S.; Attaran, M.; Celik, B.G. Digital twins and industrial internet of things: Uncovering operational intelligence in industry 4.0. *Decis. Anal. J.* **2024**, *10*, 100398. [CrossRef]
105. Richter, M.A.; Hagenmaier, M.; Bandte, O.; Parida, V.; Wincent, J. Smart cities, urban mobility and autonomous vehicles: How different cities needs different sustainable investment strategies. *Technol. Forecast. Soc. Chang.* **2022**, *184*, 121857. [CrossRef]

106. Xia, H.; Liu, Z.; Efremochkina, M.; Liu, X.; Lin, C. Study on city digital twin technologies for sustainable smart city design: A review and bibliometric analysis of geographic information system and building information modeling integration. *Sustain. Cities Soc.* **2022**, *84*, 104009. [CrossRef]
107. Yeh, A.G. From urban modelling, GIS, the digital, intelligent, and the smart city to the digital twin city with AI. *Environ. Plan. B Urban Anal. City Sci.* **2024**, *51*, 1085–1088. [CrossRef]
108. Riedelsheimer, T.; Neugebauer, S.; Lindow, K. Progress for life cycle sustainability assessment by means of digital lifecycle twins—A taxonomy. In *EcoDesign and Sustainability II*; Springer: Singapore, 2021; pp. 329–345.
109. Tetiranont, S.; Sadakorn, W.; Sreshthaputra, A.; Aniwattanapong, D.; Prasittisopin, L. Beyond the hype: Quantifying the impact of VR visualization on client preferences and price perception in AEC client pitching. *Arch. Sci. Rev.* **2024**, 1–16. [CrossRef]
110. Waqar, A.; Othman, I.; Almujibah, H.; Khan, M.B.; Alotaibi, S.; Elhassan, A.A. Factors influencing adoption of digital twin advanced technologies for smart city development: Evidence from Malaysia. *Buildings* **2023**, *13*, 775. [CrossRef]
111. Omrany, H.; Al-Obaidi, K.M. Application of digital twin technology for urban heat island mitigation: Review and conceptual framework. *Smart Sustain. Built Environ.* 2024; *ahead-of-print*. [CrossRef]
112. Kumalasari, D.; Koeva, M.; Vahdatikhaki, F.; Petrova Antonova, D.; Kuffer, M. Planning walkable cities: Generative design approach towards digital twin implementation. *Remote Sens.* **2023**, *15*, 1088. [CrossRef]
113. Lee, J.; Babcock, J.; Pham, T.S.; Bui, T.H.; Kang, M. Smart city as a social transition towards inclusive development through technology: A tale of four smart cities. *Int. J. Urban Sci.* **2023**, *27* (Suppl. S1), 75–100. [CrossRef]
114. Prasittisopin, L.; Kitkuakul, P.; Chotchakornpant, K.; Rugkhapan, N.T. Implementing the sister city policy: Perspectives from Thailand. *Nakhara J. Environ. Des. Plan.* **2024**, *23*, 413. [CrossRef]
115. Cui, J.; Ren, L.; Mai, J.; Zheng, P.; Zhang, L. 3D printing in the context of cloud manufacturing. *Robot. Comput. Integr. Manuf.* **2022**, *74*, 102256. [CrossRef]
116. Kanthimathi, T.; Rathika, N.; Fathima, A.J.; Rajesh, K.S.; Srinivasan, S.; Thamizhamuthu, R. Robotic 3D Printing for Customized Industrial Components: IoT and AI-Enabled Innovation. In Proceedings of the 2024 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 18–19 January 2024; pp. 509–513.
117. Yang, F.; Gu, S. Industry 4.0, a revolution that requires technology and national strategies. *Complex Intell. Syst.* **2021**, *7*, 1311–1325. [CrossRef]
118. Belli, L.; Cilfone, A.; Davoli, L.; Ferrari, G.; Adorni, P.; Di Nocera, F.; Dall'Olio, A.; Pellegrini, C.; Mordacci, M.; Bertolotti, E. IoT-enabled smart sustainable cities: Challenges and approaches. *Smart Cities* **2020**, *3*, 1039–1071. [CrossRef]
119. Kang, K.; Tan, B.Q.; Zhong, R.Y. Cloud-based 3D printing service allocation models for mass customization. *Int. J. Adv. Manuf. Technol.* **2023**, *126*, 2129–2145. [CrossRef]
120. Rimmer, M. Automating Fab Cities: 3D Printing and Urban Renewal. In *Automating Cities: Design, Construction, Operation and Future Impact*; Wang, B.T., Wang, C.M., Eds.; Springer: Singapore, 2021; pp. 255–272.
121. Saad Alotaibi, B.; Ibrahim Shema, A.; Umar Ibrahim, A.; Awad Abuhussain, M.; Abdulmalik, H.; Aminu Dodo, Y.; Atakara, C. Assimilation of 3D printing, Artificial Intelligence (AI) and Internet of Things (IoT) for the construction of eco-friendly intelligent homes: An explorative review. *Heliyon* **2024**, *10*, e36846. [CrossRef] [PubMed]
122. Alavi, A.H.; Feng, M.; Jiao, P.; Sharif-Khodaei, Z. *The Rise of Smart Cities: Advanced Structural Sensing and Monitoring Systems*; Butterworth-Heinemann: Oxford, UK, 2022.
123. Mohamed, A.S.Y. 3D Printing for Smart Urbanism: Towards City Livability. In *Research and Innovation Forum 2023*; Visvizi, A., Troisi, O., Corvello, V., Eds.; Springer International Publishing: Cham, Switzerland, 2024; pp. 205–217.
124. de Souza, E.A.; Borges, P.H.R.; Stengel, T.; Nematollahi, B.; Bos, F.P. 3D printed sustainable low-cost materials for construction of affordable social housing in Brazil: Potential, challenges, and research needs. *J. Build. Eng.* **2024**, *87*, 108985. [CrossRef]
125. Prashar, G.; Vasudev, H.; Bhuddhi, D. Additive manufacturing: Expanding 3D printing horizon in industry 4.0. *Int. J. Interact. Des. Manuf.* **2023**, *17*, 2221–2235. [CrossRef]
126. Mehrpouya, M.; Dehghanghadikolaei, A.; Fotovvati, B.; Vosooghnia, A.; Emamian, S.S.; Gisario, A. The potential of additive manufacturing in the smart factory industrial 4.0: A review. *Appl. Sci.* **2019**, *9*, 3865. [CrossRef]
127. Nielsen, K.; Brunoe, T.D.; Jensen, K.N.; Andersen, A.-L. Utilization of Mass Customization in Construction and Building Industry. In *Managing Complexity*; Bellemare, J., Carrier, S., Nielsen, K., Piller, F.T., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 115–125.
128. Placino, P.; Rugkhapan, N.T. Making visible concrete's shadow places: Mixing environmental concerns and social inequalities into building materials. *Environ. Urban.* **2024**, *36*, 195–213. [CrossRef]
129. Syed, A.S.; Sierra-Sosa, D.; Kumar, A.; Elmaghraby, A. IoT in smart cities: A survey of technologies, practices and challenges. *Smart Cities* **2021**, *4*, 429–475. [CrossRef]
130. Kumar, A. A novel framework for waste management in smart city transformation with industry 4.0 technologies. *Res. Glob.* **2024**, *9*, 100234. [CrossRef]
131. Goswami, L.; Singh, D. A Review on smart city using IOT. *Int. J. Innov. Res. Comput. Sci. Technol.* **2022**, *10*, 88–91. [CrossRef]
132. Demir, E.; Eyers, D.; Huang, Y. Competing through the last mile: Strategic 3D printing in a city logistics context. *Comput. Oper. Res.* **2021**, *131*, 105248. [CrossRef]
133. Olsson, N.O.E.; Arica, E.; Woods, R.; Madrid, J.A. Industry 4.0 in a project context: Introducing 3D printing in construction projects. *Proj. Leadersh. Soc.* **2021**, *2*, 100033. [CrossRef]

134. Sepasgozar, S.M.E.; Shi, A.; Yang, L.; Shirowzhan, S.; Edwards, D.J. Additive manufacturing applications for industry 4.0: A systematic critical review. *Buildings* **2020**, *10*, 231. [CrossRef]
135. Balasubramanian, S.; Shukla, V.; Islam, N.; Manghat, S. Construction industry 4.0 and sustainability: An enabling framework. *IEEE Trans. Eng. Manag.* **2021**, *71*, 1–19. [CrossRef]
136. Tsaramirsis, G.; Kantaros, A.; Al-Darraji, I.; Piromalis, D.; Apostolopoulos, C.; Pavlopoulou, A.; Alrammal, M.; Ismail, Z.; Buhari, S.M.; Stojmenovic, M. A modern approach towards an industry 4.0 model: From driving technologies to management. *J. Sens.* **2022**, *2022*, 5023011. [CrossRef]
137. Adel, A. The convergence of intelligent tutoring, robotics, and IoT in smart education for the transition from industry 4.0 to 5.0. *Smart Cities* **2024**, *7*, 325–369. [CrossRef]
138. Su, M.; Song, Y. Printable smart materials and devices: Strategies and applications. *Chem. Rev.* **2021**, *122*, 5144–5164. [CrossRef]
139. Prasittisopin, L.; Jiramarootapong, P.; Pongpaisanseree, K.; Snguanyat, C. Lean manufacturing and thermal enhancement of single-layer wall with an additive manufacturing (AM) structure. *ZKG Int.* **2019**, *4*, 64–74.
140. Yan, D.; Wang, Z.; Zhang, Z. Stimuli-responsive crystalline smart materials: From rational design and fabrication to applications. *Acc. Chem. Res.* **2022**, *55*, 1047–1058. [CrossRef] [PubMed]
141. Sobczyk, M.; Wiesenhütter, S.; Noennig, J.R.; Wallmersperger, T. Smart materials in architecture for actuator and sensor applications: A review. *J. Intell. Mater. Syst. Struct.* **2022**, *33*, 379–399. [CrossRef]
142. Liu, Y.; Zhong, Y.; Wang, C. Recent advances in self-actuation and self-sensing materials: State of the art and future perspectives. *Talanta* **2020**, *212*, 120808. [CrossRef]
143. Xia, X.; Spadaccini, C.M.; Greer, J.R. Responsive materials architected in space and time. *Nat. Rev. Mater.* **2022**, *7*, 683–701. [CrossRef]
144. Kwan, K.W.; Ngan, A.H.W. A high-performing, visible-light-driven actuating material responsive to ultralow light intensities. *Adv. Mater. Technol.* **2019**, *4*, 1900746. [CrossRef]
145. Zhang, G.; Yang, Y.; Yang, G. Smart supply chain management in industry 4.0: The review, research agenda and strategies in North America. *Ann. Oper. Res.* **2023**, *322*, 1075–1117. [CrossRef] [PubMed]
146. Kwon, H.; Soni, P.; Saeedi, A.; Shahverdi, M.; Dillenburger, B. 3D Printing and Shape Memory Alloys. In Proceedings of the 42nd Annual Conference of the Association for Computer Aided Design in Architecture, ACADIA 2022, Philadelphia, PA, USA, 27–29 October 2022; pp. 76–89.
147. Nicolay, P.; Schlögl, S.; Thaler, S.M.; Humbert, C.; Filipitsch, B. Smart Materials for Green (er) Cities, a Short Review. *Appl. Sci.* **2023**, *13*, 9289. [CrossRef]
148. Srivastava, R.; Alsamhi, S.H.; Murray, N.; Devine, D. Shape memory alloy-based wearables: A review, and conceptual frameworks on HCI and HRI in Industry 4.0. *Sensors* **2022**, *22*, 6802. [CrossRef]
149. Niazy, D.; Ashraf, M.; Bodaghi, M.; Zolfagharian, A. Resilient city perspective: 4D printing in art, architecture and construction. *Mater. Today Sustain.* **2024**, *26*, 100708. [CrossRef]
150. Chatterjee, S.; Xu, J.; Huynh, T.; Abhishek, K.; Kumari, S.; Behera, A. Performance of smart alloys in manufacturing processes during subtractive and additive Manufacturing: A short review on SMA and metal alloys. In *Smart 3D Nanoprinting*; CRC Press: Boca Raton, FL, USA, 2022; pp. 267–280.
151. Pecunia, V.; Occhipinti, L.G.; Hoye, R.L. Emerging indoor photovoltaic technologies for sustainable internet of things. *Adv. Energy Mater.* **2021**, *11*, 2100698. [CrossRef]
152. Don Chua, W.F.; Lim, C.L.; Koh, Y.Y.; Kok, C.L. A novel IoT photovoltaic-powered water irrigation control and monitoring system for sustainable city farming. *Electronics* **2024**, *13*, 676. [CrossRef]
153. Maraveas, C.; Loukatos, D.; Bartzanas, T.; Arvanitis, K.G.; Uijterwaal, J.F. Smart and solar greenhouse covers: Recent developments and future perspectives. *Front. Energy Res.* **2021**, *9*, 783587. [CrossRef]
154. Liu, L.; Guo, X.; Lee, C. Promoting smart cities into the 5G era with multi-field Internet of Things (IoT) applications powered with advanced mechanical energy harvesters. *Nano Energy* **2021**, *88*, 106304. [CrossRef]
155. Choudhary, P.; Bhargava, L.; Suhag, A.K.; Choudhary, M.; Singh, S. An era of internet of things leads to smart cities initiatives towards urbanization. In *Digital Cities Roadmap: IoT-Based Architecture and Sustainable Buildings*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2021; pp. 319–350.
156. Deng, Z.; Li, W.; Dong, W.; Sun, Z.; Kodikara, J.; Sheng, D. Multifunctional asphalt concrete pavement toward smart transport infrastructure: Design, performance and perspective. *Compos. Part B Eng.* **2023**, *265*, 110937. [CrossRef]
157. Andreu, A.; Lee, H.; Kang, J.; Yoon, Y.J. Self-healing materials for 3D printing. *Adv. Funct. Mater.* **2024**, *34*, 2315046. [CrossRef]
158. Ryan, K.R.; Down, M.P.; Banks, C.E. Future of additive manufacturing: Overview of 4D and 3D printed smart and advanced materials and their applications. *Chem. Eng. J.* **2021**, *403*, 126162. [CrossRef]
159. Miao, J.-T.; Ge, M.; Wu, Y.; Peng, S.; Zheng, L.; Chou, T.Y.; Wu, L. 3D printing of sacrificial thermosetting mold for building near-infrared irradiation induced self-healable 3D smart structures. *Chem. Eng. J.* **2022**, *427*, 131580. [CrossRef]
160. Szczucka-Lasota, B.; Węgrzyn, T.; Silva, A.P.; Jurek, A. AHSS—Construction material used in smart cities. *Smart Cities* **2023**, *6*, 1132–1151. [CrossRef]
161. Abedi, M.; Al-Jabri, K.; Han, B.; Fangueiro, R.; Lourenço, P.B.; Correia, A.G. Advancing infrastructure resilience: A polymeric composite reinforcement grid with self-sensing and self-heating capabilities. *Constr. Build. Mater.* **2024**, *435*, 136730. [CrossRef]

162. Chen, G.; Wang, K.; Yang, J.; Huang, J.; Chen, Z.; Zheng, J.; Wang, J.; Yang, H.; Li, S.; Miao, Y. Printable thermochromic hydrogel-based smart window for all-weather building temperature regulation in diverse climates. *Adv. Mater.* **2023**, *35*, 2211716. [CrossRef]
163. Chen, G.; Zhang, Y.; Chen, Y.; Fu, J. Energy efficient thermochromic smart windows based on polymers and metal oxides. *J. Polym. Sci.* **2024**, *62*, 229–240. [CrossRef]
164. Wang, D.; Chen, G.; Fu, J. Multifunctional thermochromic smart windows for building energy saving. *J. Mater. Chem. A* **2024**, *12*, 12960–12982. [CrossRef]
165. Supian, A.; Asyraf, M.; Syamsir, A.; Najeeb, M.; Alhayek, A.; Al-Dala'ien, R.N.; Manar, G.; Atiqah, A. Thermochromic polymer nanocomposites for the heat detection system: Recent progress on properties, applications, and challenges. *Polymers* **2024**, *16*, 1545. [CrossRef]
166. Jiang, C.; He, L.; Xuan, Q.; Liao, Y.; Dai, J.-G.; Lei, D. Phase-change VO2-based thermochromic smart windows. *Light: Sci. Appl.* **2024**, *13*, 255. [CrossRef]
167. Sekhar, M.C.; Veena, E.; Kumar, N.S.; Naidu, K.C.B.; Mallikarjuna, A.; Basha, D.B. A review on piezoelectric materials and their applications. *Cryst. Res. Technol.* **2023**, *58*, 2200130. [CrossRef]
168. Yuan, X.; Mai, Z.; Li, Z.; Yu, Z.; Ci, P.; Dong, S. A 3D-printing approach toward flexible piezoelectronics with function diversity. *Mater. Today* **2023**, *69*, 160–192. [CrossRef]
169. Pawar, O.; Lim, S. 3D-Printed piezoelectric nanogenerator with aligned graphitic carbon nitrate nanosheets for enhancing piezoelectric performance. *J. Colloid Interface Sci.* **2024**, *654*, 868–877. [CrossRef] [PubMed]
170. Yan, M.; Li, H.; Liu, S.; Xiao, Z.; Yuan, X.; Zhai, D.; Zhou, K.; Bowen, C.R.; Zhang, Y.; Zhang, D. 3D-printed flexible PVDF-TrFE composites with aligned BCZT nanowires and interdigital electrodes for piezoelectric nanogenerator applications. *ACS Appl. Polym. Mater.* **2023**, *5*, 4879–4888. [CrossRef]
171. Li, T.; Lee, P.S. Piezoelectric energy harvesting technology: From materials, structures, to applications. *Small Struct.* **2022**, *3*, 2100128. [CrossRef]
172. Huang, M.; Zhu, M.; Feng, X.; Zhang, Z.; Tang, T.; Guo, X.; Chen, T.; Liu, H.; Sun, L.; Lee, C. Intelligent cubic-designed piezoelectric node (iCUPE) with simultaneous sensing and energy harvesting ability toward self-sustained artificial intelligence of things (AIoT). *ACS Nano* **2023**, *17*, 6435–6451. [CrossRef] [PubMed]
173. Ravikumar, C. IoT applications powered by piezoelectric vibration energy harvesting device. In *Information and Software Technologies*; Springer: Cham, Switzerland, 2022; pp. 171–182.
174. Tuloup, C.; Harizi, W.; Aboura, Z.; Meyer, Y. Integration of piezoelectric transducers (PZT and PVDF) within polymer-matrix composites for structural health monitoring applications: New success and challenges. *Int. J. Smart Nano Mater.* **2020**, *11*, 343–369. [CrossRef]
175. Pu, W.; Wei, F.; Yao, L.; Xie, S. A review of humidity-driven actuator: Toward high response speed and practical applications. *J. Mater. Sci.* **2022**, *57*, 12202–12235. [CrossRef]
176. Choi, A.H.; Ben-Nissan, B. *Hydrogel for Biomedical Applications: 3D/4D Printing, Self-Healing, Microrobots, and Nanogenerators*; Springer Nature: Singapore, 2024.
177. Kuang, X.; Chen, K.; Dunn, C.K.; Wu, J.; Li, V.C.; Qi, H.J. 3D printing of highly stretchable, shape-memory, and self-healing elastomer toward novel 4D printing. *ACS Appl. Mater. Interfaces* **2018**, *10*, 7381–7388. [CrossRef]
178. Lu, B.; Li, H.; Wong, T.N.; Qian, S. Development of a functional cementitious mixture with expanded graphite for automated spray construction. *J. Mater. Civ. Eng.* **2023**, *35*, 04023226. [CrossRef]
179. Wang, X.; Li, Z.; Han, B.; Han, B.; Yu, X.; Zeng, S.; Ou, J. Intelligent concrete with self-x capabilities for smart cities. *J. Smart Cities* **2019**, *2*, 1–39. [CrossRef]
180. Lawongkerd, J.; Vichai, K.; Thamniap, B.; Prasittisopin, L.; Saensuk, O.; Keawsawasvong, S. A study of thermoelectric energy harvesting on asphalt concrete pavement. *Transp. Infrastruct. Geotechnol.* **2024**, *11*, 1448–1461. [CrossRef]
181. Yang, S.; Zhu, H.; Yang, X.; Tan, Q.; Chen, Y. Fabrication and performance evaluation of high-performance SBS-modified asphalt through secondary modification with aminated graphene oxide. *J. Mater. Civ. Eng.* **2024**, *36*, 04024417. [CrossRef]
182. Win, T.T.; Prasittisopin, L.; Jongvivatsakul, P.; Likitlersuang, S. Investigating the synergistic effect of graphene nanoplatelets and fly ash on the mechanical properties and microstructure of calcium aluminate cement composites. *J. Build. Eng.* **2023**, *78*, 107710. [CrossRef]
183. Salah, H.A.; Mutalib, A.A.; Kaish, A.; Syamsir, A.; Algaifi, H.A. Development of ultra-high-performance silica fume-based mortar incorporating graphene nanoplatelets for 3-dimensional concrete printing application. *Buildings* **2023**, *13*, 1949. [CrossRef]
184. Schulte, J.; Jiang, Z.; Sevim, O.; Ozbulut, O.E. Graphene-reinforced cement composites for smart infrastructure systems. In *Rise Smart Cities*; Butterworth-Heinemann: Oxford, UK, 2022; pp. 79–114.
185. Jiang, X.; Lu, D.; Yin, B.; Leng, Z. Advancing carbon nanomaterials-engineered self-sensing cement composites for structural health monitoring: A state-of-the-art review. *J. Build. Eng.* **2024**, *87*, 109129. [CrossRef]
186. D'Alessandro, A.; Birgin, H.B.; Ubertini, F. Advanced monitoring of structures and infrastructures through smart composite sensors and systems. In *Civil Structural Health Monitoring*; Springer: Cham, Switzerland, 2021; pp. 485–498.
187. Win, T.T.; Raengthon, N.; Prasittisopin, L. Advanced cement composites: Investigating the role of graphene quantum dots in improving thermal and mechanical performance. *J. Build. Eng.* **2024**, *96*, 110556. [CrossRef]

188. Win, T.T.; Prasittisopin, L.; Nganglumpoon, R.; Pinthong, P.; Watmanee, S.; Tolek, W.; Panpranot, J. Chemo-physical mechanisms of high-strength cement composites with suprastructure of graphene quantum dots. *Clean. Mater.* **2024**, *11*, 100229. [CrossRef]
189. Raj, A.; Yamkasikorn, P.; Wangtawesap, R.; Win, T.T.; Ngamkhanong, C.; Jongvivatsakul, P.; Prasittisopin, L.; Panpranot, J.; Kaewunruen, S. Effect of Graphene Quantum Dots (GQDs) on the mechanical, dynamic, and durability properties of concrete. *Constr. Build. Mater.* **2024**, *441*, 137597. [CrossRef]
190. Win, T.T.; Prasittisopin, L.; Nganglumpoon, R.; Pinthong, P.; Watmanee, S.; Tolek, W.; Panpranot, J. Innovative GQDs and supra-GQDs assemblies for developing high strength and conductive cement composites. *Constr. Build. Mater.* **2024**, *421*, 135693. [CrossRef]
191. Ismail, K.A.; Lino, F.A.; Henríquez, J.R.; Teggar, M.; Laouer, A.; Arici, M.; Benhorma, A.; Rodríguez, D. Enhancement techniques for the reduction of heating and cooling loads in buildings: A review. *J. Energy Power Technol.* **2023**, *5*, 031. [CrossRef]
192. Block, A.B.; Palou, J.E.; Courtant, M.; Virtuani, A.; Cattaneo, G.; Roten, M.; Li, H.; Despeisse, M.; Hessler-Wyser, A.; Desai, U. Colouring solutions for building integrated photovoltaic modules: A review. *Energy Build.* **2024**, *314*, 114253. [CrossRef]
193. Sadakorn, W.; Tetiranont, S.; Prasittisopin, L.; Kaewunruen, S. Assessing thermal comfort in hot and humid (tropical) climates: Urban outdoor and semi-outdoor conditions in waiting areas of railway stations. *Build. Environ.* **2024**, *267*, 112240. [CrossRef]
194. Del Pero, C.; Leonforte, F.; Aste, N. Building-integrated photovoltaics in existing buildings: A novel PV roofing system. *Buildings* **2024**, *14*, 2270. [CrossRef]
195. Ciampi, G.; Spanodimitriou, Y.; Scorpio, M.; Rosato, A.; Sibilio, S. Energy performances assessment of extruded and 3d printed polymers integrated into building envelopes for a south Italian case study. *Buildings* **2021**, *11*, 141. [CrossRef]
196. Raja, S.; Mustafa, M.A.; Ghadir, G.K.; Al-Tmimi, H.M.; Alani, Z.K.; Rusho, M.A.; Rajeswari, N. Unlocking the potential of polymer 3D printed electronics: Challenges and solutions. *Appl. Chem. Eng.* **2024**, *7*, 3877. [CrossRef]
197. Nguyen, P.Q.; Courchesne, N.M.D.; Duraj-Thatte, A.; Praveschotinunt, P.; Joshi, N.S. Engineered living materials: Prospects and challenges for using biological systems to direct the assembly of smart materials. *Adv. Mater.* **2018**, *30*, 1704847. [CrossRef]
198. Gao, W.; Zhang, Y.; Ramanujan, D.; Ramani, K.; Chen, Y.; Williams, C.B.; Wang, C.C.; Shin, Y.C.; Zhang, S.; Zavattieri, P.D. The status, challenges, and future of additive manufacturing in engineering. *Comput. Aided Des.* **2015**, *69*, 65–89. [CrossRef]
199. Son, J.-h.; Kim, H.; Choi, Y.; Lee, H. 3D printed energy devices: Generation, conversion, and storage. *Microsyst. Nanoeng.* **2024**, *10*, 93. [CrossRef]
200. Li, H.; Liang, J. Recent development of printed micro-supercapacitors: Printable materials, printing technologies, and perspectives. *Adv. Mater.* **2020**, *32*, 1805864. [CrossRef]
201. Trivedi, D.; Rahn, C.D.; Kier, W.M.; Walker, I.D. Soft robotics: Biological inspiration, state of the art, and future research. *Appl. Bionics Biomech.* **2008**, *5*, 99–117. [CrossRef]
202. Seong, M.; Sun, K.; Kim, S.; Kwon, H.; Lee, S.W.; Veerla, S.C.; Kang, D.K.; Kim, J.; Kondaveeti, S.; Tawfik, S.M.; et al. Multifunctional Magnetic Muscles for Soft Robotics. *Nat. Commun.* **2024**, *15*, 7929. [CrossRef] [PubMed]
203. Wang, Z.; Wu, Y.; Zhu, B.; Chen, Q.; Wang, L.; Zhao, Y.; Sun, D.; Zheng, J.; Wu, D. A magnetic soft robot with multimodal sensing capability by multimaterial direct ink writing. *Addit. Manuf.* **2023**, *61*, 103320. [CrossRef]
204. Sun, J.; Tighe, B.; Liu, Y.; Zhao, J. Twisted-and-coiled actuators with free strokes enable soft robots with programmable motions. *Soft Robot.* **2021**, *8*, 213–225. [CrossRef]
205. Zhao, R.; Dai, H.; Yao, H. Liquid-metal magnetic soft robot with reprogrammable magnetization and stiffness. *IEEE Robot. Autom. Lett.* **2022**, *7*, 4535–4541. [CrossRef]
206. Zhang, Y.; Li, P.; Quan, J.; Li, L.; Zhang, G.; Zhou, D. Progress, challenges, and prospects of soft robotics for space applications. *Adv. Intell. Syst.* **2023**, *5*, 2200071. [CrossRef]
207. Liao, J.; Majidi, C.; Sitti, M. Liquid metal actuators: A comparative analysis of surface tension controlled actuation. *Adv. Mater.* **2024**, *36*, 2300560. [CrossRef] [PubMed]
208. Wallin, T.J.; Pikul, J.; Shepherd, R.F. 3D printing of soft robotic systems. *Nat. Rev. Mater.* **2018**, *3*, 84–100. [CrossRef]
209. Sadeghi, A.; Mondini, A.; Mazzolai, B. Toward self-growing soft robots inspired by plant roots and based on additive manufacturing technologies. *Soft Robot.* **2017**, *4*, 211–223. [CrossRef]
210. Bier, H.; Hidding, A.; Latour, M.; Oskam, P.; Alavi, H.; Külekci, A. Design-to-robotic-production and-operation for activating bio-cyber-physical environments. In *Disruptive Technologies: The Convergence of New Paradigms in Architecture*; Springer: Cham, Switzerland, 2023; pp. 45–57.
211. Sadeghi, A.; Del Dottore, E.; Mondini, A.; Mazzolai, B. Passive morphological adaptation for obstacle avoidance in a self-growing robot produced by additive manufacturing. *Soft Robot.* **2020**, *7*, 85–94. [CrossRef]
212. Fernandez, S.V.; Sadat, D.; Tasnim, F.; Acosta, D.; Schwendeman, L.; Shahsavari, S.; Dagdeviren, C. Ubiquitous conformable systems for imperceptible computing. *Foresight* **2022**, *24*, 75–98. [CrossRef]
213. Block, A.E. HuggieBot: An Interactive Hugging Robot with Visual and Haptic Perception. Ph.D. Thesis, ETH Zurich, Zürich, Switzerland, 2021.
214. Hang, K.; Lyu, X.; Song, H.; Stork, J.A.; Dollar, A.M.; Kragic, D.; Zhang, F. Perching and resting—A paradigm for UAV maneuvering with modularized landing gears. *Sci. Robot.* **2019**, *4*, eaau6637. [CrossRef] [PubMed]
215. Tawk, C.; Zhou, H.; Sariyildiz, E.; Panhuis, M.I.H.; Spinks, G.M.; Alici, G. Design, modeling, and control of a 3D printed monolithic soft robotic finger with embedded pneumatic sensing chambers. *IEEE/ASME Trans. Mechatron.* **2020**, *26*, 876–887. [CrossRef]

216. Tawk, C.; Sariyildiz, E.; Alici, G. Force control of a 3D printed soft gripper with built-in pneumatic touch sensing chambers. *Soft Robot.* **2022**, *9*, 970–980. [CrossRef]
217. Roy, M.; Roy, A. The rise of interdisciplinarity in engineering education in the era of industry 4.0: Implications for management practice. *IEEE Eng. Manag. Rev.* **2021**, *49*, 56–70. [CrossRef]
218. Khanna, J.; AL-Rawi, M.; Hartley, C. Development of innovative cross-disciplinary engineering showcase. In Proceedings of the MaDE2020: Synergies in New Zealand Manufacturing, Design and Entrepreneurship, Auckland, New Zealand, 7–8 December 2020.
219. Treutler, K.; Wesling, V. The current state of research of wire arc additive manufacturing (WAAM): A review. *Appl. Sci.* **2021**, *11*, 8619. [CrossRef]
220. Wei, H.L.; Bhadeshia, H.K.D.H.; David, S.A.; DebRoy, T. Harnessing the scientific synergy of welding and additive manufacturing. *Sci. Technol. Weld. Join.* **2019**, *24*, 361–366. [CrossRef]
221. Dilibal, S.; Nohut, S.; Kurtoglu, C.; Owusu-Danquah, J. Data-driven generative design integrated with hybrid additive subtractive manufacturing (HASM) for smart cities. In *Data-Driven Mining, Learning and Analytics for Secured Smart Cities: Trends and Advances*; Springer: Cham, Switzerland, 2021; pp. 205–228.
222. Casas-Bocanegra, D.; Gomez-Vargas, D.; Pinto-Bernal, M.J.; Maldonado, J.; Munera, M.; Villa-Moreno, A.; Stoelen, M.F.; Belpaeme, T.; Cifuentes, C.A. An open-source social robot based on compliant soft robotics for therapy with children with ASD. *Actuators* **2020**, *9*, 91. [CrossRef]
223. Xi, W. Enhancing awareness of urban greenbelts through the integration of soft robots and cultural storytelling: A more-than-human design approach. Master's Thesis, Leiden University, Leiden, The Netherlands, 2023.
224. Xia, C.; Pan, Z.; Polden, J.; Li, H.; Xu, Y.; Chen, S.; Zhang, Y. A review on wire arc additive manufacturing: Monitoring, control and a framework of automated system. *J. Manuf. Syst.* **2020**, *57*, 31–45. [CrossRef]
225. Fidan, I.; Huseynov, O.; Ali, M.A.; Alkunte, S.; Rajeshirke, M.; Gupta, A.; Hasanov, S.; Tantawi, K.; Yasa, E.; Yilmaz, O. Recent inventions in additive manufacturing: Holistic review. *Inventions* **2023**, *8*, 103. [CrossRef]
226. Costanzi, C.B.; Waldschmitt, B.; Knaack, U.; Lange, J. Transforming the construction industry through wire arc additive manufacturing. In *Coding Architecture: Designing Toolkits, Workflows, Industry*; Springer: Cham, Switzerland, 2023; pp. 213–238.
227. Chadha, K.; Dubor, A.; Cabay, E.; Tayoun, Y.; Naldoni, L.; Moretti, M. Additive manufacturing for the circular built environment: Towards circular construction with earth-based materials. In *A Circular Built Environment in the Digital Age*; Springer International Publishing: Cham, Switzerland, 2024; pp. 111–128.
228. Dörrie, R.; Laghi, V.; Arrè, L.; Kienbaum, G.; Babovic, N.; Hack, N.; Kloft, H. Combined additive manufacturing techniques for adaptive coastline protection structures. *Buildings* **2022**, *12*, 1806. [CrossRef]
229. Fu, H.; Kaewunruen, S. State-of-the-art review on additive manufacturing technology in railway infrastructure systems. *J. Compos. Sci.* **2022**, *6*, 7. [CrossRef]
230. Rane, N.; Choudhary, S.; Rane, J. *4D/5D/6D Printing Technology in the Architecture, Engineering, and Construction (AEC) Industry: Applications, Challenges, and Future Advancements*; Elsevier: Amsterdam, The Netherlands, 2023.
231. Khanpara, P.; Tanwar, S. Additive manufacturing: Concepts and technologies. In *A Roadmap to Industry 4.0: Smart Production, Sharp Business and Sustainable Development*; Springer: Cham, Switzerland, 2020; pp. 171–185.
232. Laghi, V.; Palermo, M.; Gasparini, G.; Trombetti, T. Wire and arc additive manufacturing: Wire and arc additive manufacturing technology: A research perspective. In *Additive Manufacturing for Construction*; Emerald Group Publishing Ltd.: Leeds, UK, 2023; pp. 141–159.
233. Nyamuchiwa, K.; Palad, R.; Panlican, J.; Tian, Y.; Aranas, C., Jr. Recent progress in hybrid additive manufacturing of metallic materials. *Appl. Sci.* **2023**, *13*, 8383. [CrossRef]
234. Helm, J.M.; Swiergosz, A.M.; Haeberle, H.S.; Karnuta, J.M.; Schaffer, J.L.; Krebs, V.E.; Spitzer, A.I.; Ramkumar, P.N. Machine learning and artificial intelligence: Definitions, applications, and future directions. *Curr. Rev. Musculoskelet. Med.* **2020**, *13*, 69–76. [CrossRef]
235. Mahesh, B. Machine learning algorithms-a review. *Int. J. Sci. Res.* **2020**, *9*, 381–386. [CrossRef]
236. Mondal, K.; Martinez, O.; Jain, P. Advanced manufacturing and digital twin technology for nuclear energy. *Front. Energy Res.* **2024**, *12*, 1339836. [CrossRef]
237. Brenner, M.; Langenberg, S.; Angermann, K.; Meier, H.-R. *High-Tech Heritage:(Im) Permanence of Innovative Architecture*; Birkhäuser: Basel, Switzerland, 2024.
238. Nguyen, P.D.; Nguyen, T.Q.; Tao, Q.; Vogel, F.; Nguyen-Xuan, H. A data-driven machine learning approach for the 3D printing process optimisation. *Virtual Phys. Prototyp.* **2022**, *17*, 768–786. [CrossRef]
239. Chaturvedi, M.; Vendan, S.A. Data-Driven Models in Machine Learning: An Enabler of Smart Manufacturing. In *Big Data Analytics in Smart Manufacturing: Principles and Practices*; CRC Press: Boca Raton, FL, USA, 2022; pp. 35–68.
240. Baduge, S.K.; Thilakarathna, S.; Perera, J.S.; Arashpour, M.; Sharafi, P.; Teodosio, B.; Shringi, A.; Mendis, P. Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. *Autom. Constr.* **2022**, *141*, 104440. [CrossRef]
241. Sun, X.; Zhou, K.; Demoly, F.; Zhao, R.R.; Qi, H.J. Perspective: Machine learning in design for 3D/4D printing. *J. Appl. Mech.* **2024**, *91*, 030801. [CrossRef]

242. Chigilipalli, B.K.; Karri, T.; Chetti, S.N.; Bhiogade, G.; Kottala, R.K.; Cheepu, M. A review on recent trends and applications of IoT in additive manufacturing. *Appl. Syst. Innov.* **2023**, *6*, 50. [CrossRef]
243. Talaat, F.M.; Hassan, E. Artificial intelligence in 3D printing. In *Enabling Machine Learning Applications in Data Science*; Springer: Singapore, 2021; pp. 77–88.
244. Prasittisopin, L.; Tuvayanond, W. Machine Learning for Strength Prediction of Ready-Mix Concretes Containing Chemical and Mineral Admixtures and Cured at Different Temperatures. In Proceedings of the 6th International Conference on Civil Engineering and Architecture, Bali, Indonesia, 16–18 December 2023; Springer: Singapore, 2023; pp. 242–249.
245. Tuvayanond, W.; Kamchoom, V.; Prasittisopin, L. Efficient machine learning for strength prediction of ready-mix concrete production (prolonged mixing). *Constr. Innov.* 2024; *ahead-of-print*. [CrossRef]
246. Mehta, P.; Mujawar, M.; LaFrance, S.; Bernadin, S.; Ewing, D.; Bhansali, S. Sensor-based and computational methods for error detection and correction in 3D printing. *ECS Sens. Plus* **2024**, *3*, 030602. [CrossRef]
247. Pech, M.; Vrchota, J.; Bednář, J. Predictive maintenance and intelligent sensors in smart factory. *Sensors* **2021**, *21*, 1470. [CrossRef] [PubMed]
248. Prasittisopin, L.; Trejo, D. Performance characteristics of blended cementitious systems incorporating chemically transformed rice husk Ash. *Adv. Civ. Eng. Mater.* **2017**, *6*, 17–35. [CrossRef]
249. Sereewatthanawut, I.; Panwisawas, C.; Ngamkhanong, C.; Prasittisopin, L. Effects of extended mixing processes on fresh, hardened and durable properties of cement systems incorporating fly ash. *Sci. Rep.* **2023**, *13*, 6091. [CrossRef] [PubMed]
250. Win, T.T.; Wattanapornprom, R.; Prasittisopin, L.; Pansuk, W.; Pheinsusom, P. Investigation of fineness and calcium-oxide content in fly ash from ASEAN region on properties and durability of cement–fly ash system. *Eng. J.* **2022**, *26*, 77–90. [CrossRef]
251. Sampedro, G.A.R.; Putra, M.A.P.; Abisado, M. 3D-AmplifAI: An ensemble machine learning approach to digital twin fault monitoring for additive manufacturing in smart factories. *IEEE Access* **2023**, *11*, 64128–64140. [CrossRef]
252. Suvarna, M.; Büth, L.; Hejny, J.; Mennenga, M.; Li, J.; Ng, Y.T.; Herrmann, C.; Wang, X. Smart manufacturing for smart cities—Overview, insights, and future directions. *Adv. Intell. Syst.* **2020**, *2*, 2000043. [CrossRef]
253. Jena, O.P.; Bhushan, B.; Kose, U. *Machine Learning and Deep Learning in Medical Data Analytics and Healthcare Applications*; CRC Press: Boca Raton, FL, USA, 2022.
254. Zain, M.; Keawsawasvong, S.; Thongchom, C.; Sereewatthanawut, I.; Usman, M.; Prasittisopin, L. Establishing efficacy of machine learning techniques for vulnerability information of tubular buildings. *Eng. Sci.* **2023**, *27*, 1008. [CrossRef]
255. Goh, G.D.; Sing, S.L.; Yeong, W.Y. A review on machine learning in 3D printing: Applications, potential, and challenges. *Artif. Intell. Rev.* **2021**, *54*, 63–94. [CrossRef]
256. Gharehbaghi, V.R.; Noroozinejad Farsangi, E.; Noori, M.; Yang, T.Y.; Li, S.; Nguyen, A.; Málaga-Chuquitaype, C.; Gardoni, P.; Mirjalili, S. A critical review on structural health monitoring: Definitions, methods, and perspectives. *Arch. Comput. Methods Eng.* **2022**, *29*, 2209–2235. [CrossRef]
257. Ditommaso, R.; Ponzo, F.C. Identifying Damage in Structures: Definition of Thresholds to Minimize False Alarms in SHM Systems. *Buildings* **2024**, *14*, 821. [CrossRef]
258. Amezquita-Sanchez, J.P.; Adeli, H. Signal processing techniques for vibration-based health monitoring of smart structures. *Arch. Comput. Methods Eng.* **2016**, *23*, 1–15. [CrossRef]
259. Zinno, R.; Artese, S.; Clausi, G.; Magarò, F.; Meduri, S.; Miceli, A.; Venneri, A. Structural health monitoring (SHM). In *The Internet of Things for Smart Urban Ecosystems*; Springer: Cham, Switzerland, 2019; pp. 225–249.
260. Guerrieri, A.; Loscri, V.; Rovella, A.; Fortino, G. *Management of Cyber Physical Objects in the Future Internet of Things*; Springer Nature: Cham, Switzerland, 2016; pp. 31–50.
261. Figueiredo, E.; Brownjohn, J. Three decades of statistical pattern recognition paradigm for SHM of bridges. *Struct. Health Monit.* **2022**, *21*, 3018–3054. [CrossRef]
262. Vijayan, D.S.; Sivasuriyan, A.; Devarajan, P.; Krejsa, M.; Chalecki, M.; Żółtowski, M.; Kozarzewska, A.; Koda, E. Development of intelligent technologies in SHM on the innovative diagnosis in civil engineering—A comprehensive review. *Buildings* **2023**, *13*, 1903. [CrossRef]
263. Mishra, M.; Lourenço, P.B.; Ramana, G.V. Structural health monitoring of civil engineering structures by using the internet of things: A review. *J. Build. Eng.* **2022**, *48*, 103954. [CrossRef]
264. Zinno, R.; Haghshenas, S.S.; Guido, G.; Rashvand, K.; Vitale, A.; Sarhadi, A. The state of the art of artificial intelligence approaches and new technologies in structural health monitoring of bridges. *Appl. Sci.* **2022**, *13*, 97. [CrossRef]
265. Zhao, D.; Liu, X.; Meves, J.; Billings, C.; Liu, Y. 3D printed and embedded strain sensors in structural composites for loading monitoring and damage diagnostics. *J. Compos. Sci.* **2023**, *7*, 437. [CrossRef]
266. Mousavi, S.; Blanloeuil, P.; Vinoth, T.; Howard, D.; Wang, C.H. *A 3D Printed Constriction-Resistive Sensor for the Detection of Ultrasonic Waves*; Materials Research Forum LLC.: Millersville, PA, USA, 2021; pp. 272–277.
267. Kim, T.-H.; Moeinnia, H.; Kim, W.S. 3D printed vorticella-kirigami inspired sensors for structural health monitoring in internet-of-things. *Mater. Des.* **2023**, *234*, 112332. [CrossRef]
268. Fitzpatrick, D.; Billings, C.; Liu, Y. In Lightweight Composites with 3D Printed Sensors for Real-Time Damage Detection. In Proceedings of the ASME 2023 Aerospace Structures, Structural Dynamics, and Materials Conference, San Diego, CA, USA, 19–21 June 2023; American Society of Mechanical Engineers: New York, NY, USA, 2023; p. V001T01A005.

269. Zain, M.; Dackermann, U.; Prasittisopin, L. Machine learning (ML) algorithms for seismic vulnerability assessment of school buildings in high-intensity seismic zones. *Structures* **2024**, *70*, 107639. [CrossRef]
270. Zain, M.; Prasittisopin, L.; Mehmood, T.; Ngamkhanong, C.; Keawsawasvong, S.; Thongchom, C. A novel framework for effective structural vulnerability assessment of tubular structures using machine learning algorithms (GA and ANN) for hybrid simulations. *Nonlinear Eng.* **2024**, *13*, 20220365. [CrossRef]
271. Zain, M.; Ngamkhanong, C.; Kang, T.H.K.; Usman, M.; Prasittisopin, L. Modal-based fragility analysis of high-rise tubular structures: A methodology for vulnerability assessment. *Structures* **2024**, *63*, 106289. [CrossRef]
272. Hořejší, P.; Novikov, K.; Šimon, M. A smart factory in a Smart City: Virtual and augmented reality in a smart assembly line. *IEEE Access* **2020**, *8*, 94330–94340. [CrossRef]
273. Li, S.; Chen, C.-L.; Loh, K.J. Laboratory evaluation of railroad crosslevel tilt sensing using electrical time domain reflectometry. *Sensors* **2020**, *20*, 4470. [CrossRef]
274. *EN 206-1:2000*; Concrete-Part 1: Specification, Performance, Production and Conformity. British Standards Institution: London, UK, 2000.
275. Levy, M.; Salvadori, M. *Why Buildings Fall Down: How Structures Fail*; WW Norton & Company: New York, NY, USA, 2002.
276. Kaszynska, M.; Nowak, A.S. Effect of Material Quality on Life-Time Performance of Concrete Structures. In *Life-Cycle Performance of Deteriorating Structures: Assessment, Design and Management*; American Society of Civil Engineers: Reston, VA, USA, 2024; pp. 141–147.
277. Yu, G.; Wang, Y.; Hu, M.; Shi, L.; Mao, Z.; Sugumaran, V. RIOMS: An intelligent system for operation and maintenance of urban roads using spatio-temporal data in smart cities. *Future Gener. Comput. Syst.* **2021**, *115*, 583–609. [CrossRef]
278. Yamato, Y.; Fukumoto, Y.; Kumazaki, H. Proposal of real time predictive maintenance platform with 3D printer for business vehicles. *Int. J. Inf. Electron. Eng.* **2016**, *6*, 289–293. [CrossRef]
279. Birtchnell, T. 3D printing and the changing logistics of cities. In *Handbook of Urban Mobilities*; Routledge: London, UK, 2020; pp. 348–356.
280. Lipson, H.; Kurman, M. *Fabricated: The New World of 3D Printing*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
281. Bhanji, S.; Shotz, H.; Tadanki, S.; Miloudi, Y.; Warren, P. Advanced enterprise asset management systems: Improve predictive maintenance and asset performance by leveraging industry 4.0 and the internet of things (IoT). In Proceedings of the 2021 Joint Rail Conference, Virtual, 20–21 April 2021; American Society of Mechanical Engineers: New York, NY, USA, 2021; p. V001T12A002.
282. Jandyal, A.; Chaturvedi, I.; Wazir, I.; Raina, A.; Haq, M.I.U. 3D printing–a review of processes, materials and applications in industry 4.0. *Sustain. Oper. Comput.* **2022**, *3*, 33–42. [CrossRef]
283. Arriola, J.B.; van Oudheusden, A.A.; Flipsen, B.; Faludi, J. *3D Printing for Repair Guide*; TU Delft: Delft, The Netherlands, 2022; p. 48.
284. Khan, M.; McNally, C. Recent developments on low carbon 3D printing concrete: Revolutionizing construction through innovative technology. *Clean. Mater.* **2024**, *12*, 100251. [CrossRef]
285. Zaid, O.; El Ouni, M.H. Advancements in 3D printing of cementitious materials: A review of mineral additives, properties, and systematic developments. *Constr. Build. Mater.* **2024**, *427*, 136254. [CrossRef]
286. Sereewatthanawut, I.; Pansuk, W.; Pheinsusom, P.; Prasittisopin, L. Chloride-induced corrosion of a galvanized steel-embedded calcium sulfoaluminate stucco system. *J. Build. Eng.* **2021**, *44*, 103376. [CrossRef]
287. Perrot, A.; Jacquet, Y.; Caron, J.F.; Mesnil, R.; Ducoulombier, N.; De Bono, V.; Sanjayan, J.; Ramakrishnan, S.; Kloft, H.; Gosslar, J.; et al. Snapshot on 3D printing with alternative binders and materials: Earth, geopolymers, gypsum and low carbon concrete. *Cem. Concr. Res.* **2024**, *185*, 107651. [CrossRef]
288. Kumar, V. Sustainable 3D Printing–Based Smart Solution for the Maintenance of Heritage Structures. In *Sustainable Manufacturing*; CRC Press: Boca Raton, FL, USA, 2024; pp. 13–30.
289. Compare, M.; Baraldi, P.; Zio, E. Challenges to IoT-enabled predictive maintenance for industry 4.0. *IEEE Internet Things J.* **2019**, *7*, 4585–4597. [CrossRef]
290. Sampedro, G.A.; Putra, M.A.P.; Lee, J.-M.; Kim, D.-S. Industrial internet of things-based fault mitigation for smart additive manufacturing using multi-flow BiLSTM. *IEEE Access* **2023**, *11*, 99130–99142. [CrossRef]
291. Bibri, S.E.; Bibri, S.E. Data-driven smart sustainable cities: A conceptual framework for urban intelligence functions and related processes, systems, and sciences. In *Advances in the Leading Paradigms of Urbanism and their Amalgamation: Compact Cities, Eco-Cities, and Data–Driven Smart Cities*; Springer: Cham, Switzerland, 2020; pp. 143–173.
292. Gkontzis, A.F.; Kotsiantis, S.; Feretzakis, G.; Verykios, V.S. Enhancing urban resilience: Smart city data analyses, forecasts, and digital twin techniques at the neighborhood level. *Future Internet* **2024**, *16*, 47. [CrossRef]
293. Kumar, H.; Singh, M.K.; Gupta, M.; Madaan, J. Moving towards smart cities: Solutions that lead to the smart city transformation framework. *Technol. Forecast. Soc. Change* **2020**, *153*, 119281. [CrossRef]
294. Bhagia, S.; Bornani, K.; Agrawal, R.; Satlewal, A.; Ďurkovič, J.; Lagaňa, R.; Bhagia, M.; Yoo, C.G.; Zhao, X.; Kunc, V. Critical review of FDM 3D printing of PLA biocomposites filled with biomass resources, characterization, biodegradability, upcycling and opportunities for biorefineries. *Appl. Mater. Today* **2021**, *24*, 101078. [CrossRef]

295. Luque-del-Castillo, F.-d.-P.; Ladrón-de-Guevara-Muñoz, M.C.; de-Cózar-Macías, Ó.D.; Castillo-Rueda, F.J.; Pérez-García, J.; Martínez-Torres, J.-L. Construction of a Recycled Plastic Filament Winder for 3D Printing. In *Advances in Design Engineering IV*; Manchado del Val, C., Suffo Pino, M., Miralbes Buil, R., Moreno Sánchez, D., Moreno Nieto, D., Eds.; Springer Nature: Cham, Switzerland, 2024; pp. 502–513.
296. Tu, H.; Wei, Z.; Bahrami, A.; Kahla, N.B.; Ahmad, A.; Özkılıç, Y.O. Recent advancements and future trends in 3D printing concrete using waste materials. *Dev. Built Environ.* **2023**, *16*, 100187. [CrossRef]
297. Trejo, D.; Prasittisopin, L. Chemical transformation of rice husk ash morphology. *ACI Mater. J.* **2015**, *112*, 358–392. [CrossRef]
298. Krčma, M.; Škaroupka, D.; Vosynek, P.; Zikmund, T.; Kaiser, J.; Palousek, D. Use of polymer concrete for large-scale 3D printing. *Rapid Prototyp. J.* **2021**, *27*, 465–474. [CrossRef]
299. Sereewatthanawut, I.; Prasittisopin, L. Effects of accelerating and retarding agents on nucleation and crystal growth of calcium aluminate cement. *Open Ceram.* **2022**, *11*, 100290. [CrossRef]
300. Win, T.T.; Prasittisopin, L.; Jongvivatsakul, P.; Likitlersuang, S. Investigating the role of steel and polypropylene fibers for enhancing mechanical properties and microstructural performance in mitigating conversion effects in calcium aluminate cement. *Constr. Build. Mater.* **2024**, *430*, 136515. [CrossRef]
301. Robayo-Salazar, R.; Muñoz, M.A.; Vargas, A.; de Gutiérrez, R.M. Effects of incorporating bentonite, metakaolin, microsilica, and calcium carbonate on the rheological properties of portland cement-based 3D printing inks. *Constr. Build. Mater.* **2024**, *445*, 137857. [CrossRef]
302. Trejo, D.; Prasittisopin, L. Effects of mixing variables on early-age characteristics of portland cement systems. *J. Mater. in Civil Eng.* **2016**, *28*, 04016094. [CrossRef]
303. Lu, Y.; Xiao, J.; Li, Y. 3D printing recycled concrete incorporating plant fibres: A comprehensive review. *Constr. Build. Mater.* **2024**, *425*, 135951. [CrossRef]
304. Kumar, V.; Singh, R.; Ahuja, I.S. On 3D printing of electro-active PVDF-Graphene and Mn-doped ZnO nanoparticle-based composite as a self-healing repair solution for heritage structures. *Proc. Inst. Mech. Eng. Part B J. Eng. Manuf.* **2022**, *236*, 1141–1154. [CrossRef]
305. Kumar, V.; Singh, R.; Ahuja, I.S. Multi-material printing of PVDF composites: A customized solution for maintenance of heritage structures. *Proc. Inst. Mech. Eng. Part L J. Mater. Des. Appl.* **2023**, *237*, 554–564. [CrossRef]
306. Wang, Z.; Luan, C.; Liao, G.; Yao, X.; Fu, J. Mechanical and self-monitoring behaviors of 3D printing smart continuous carbon fiber-thermoplastic lattice truss sandwich structure. *Compos. Part B Eng.* **2019**, *176*, 107215. [CrossRef]
307. Zhang, P.; Arceneaux, D.J.; Liu, Z.; Nikaeen, P.; Khattab, A.; Li, G. A crack healable syntactic foam reinforced by 3D printed healing-agent based honeycomb. *Compos. Part B Eng.* **2018**, *151*, 25–34. [CrossRef]
308. Luan, C.; Yao, X.; Zhang, C.; Fu, J.; Wang, B. Integrated self-monitoring and self-healing continuous carbon fiber reinforced thermoplastic structures using dual-material three-dimensional printing technology. *Compos. Sci. Technol.* **2020**, *188*, 107986. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

urban science

MDPI

Perspective

When Cities Go Nuclear: Exploring the Applications of Nuclear Batteries Toward Energy Transformation

Sanjana Paul [1], Mikita Klimenka [1], Fabio Duarte [1,*], Carmen Crawford [2], Claire Gorman [1], Carlo Ratti [1] and Jacopo Buongiorno [2]

[1] Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; sanjanap@mit.edu (S.P.); klimenko@mit.edu (M.K.); clairego@mit.edu (C.G.); ratti@mit.edu (C.R.)
[2] Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; csleight@mit.edu (C.C.); jacopo@mit.edu (J.B.)
* Correspondence: fduarte@mit.edu

Abstract: Global society faces the pressing question of how to eliminate reliance on fossil fuels while meeting increasing energy demand. In comparison to solar and wind energy, nuclear power has been largely ignored in urban studies research. However, nuclear energy has recently regained attention through the emergence of Small Modular Reactors (SMRs), and as the stakes of decarbonization become increasingly essential. To evaluate situations in which SMRs bring value to urban energy mixes, this paper focuses on Nuclear Batteries (NBs), a specific class of SMRs, that can fit in standard shipping containers. First, we outline an evaluation framework for the use and application of NBs; second, we present use cases for NBs in real-world situations, from disaster relief to grid reinforcement; and third, we discuss the social challenges around this technology.

Keywords: nuclear energy; energy transition; urban design

1. Introduction

By 2050, global energy consumption could increase by up to 50% [1]. Energy demand will be driven by global population growth, increasing access to electricity, rising incomes, and urbanization [1]. This demand will be spatially unevenly distributed, with accelerating urbanization happening in the global south, and the industrial hinterland feeding their appetites for food, water, fuel, data, and other resources. Manufacturing, materials processing, transportation, and buildings are already responsible for over 80% of global carbon emissions [2].

Fossil fuels continue to dominate the energy market, producing more than 80% of global energy [3]. The projected growth in energy demand over the next 30 years will occur in regions heavily reliant on CO_2-emitting energy sources [1] The International Energy Agency [4] has set the imperative of transitioning to net-zero carbon emissions by 2050 to avert the worst effects of climate change.

Renewable energy sources are the fastest-growing segment of the energy sector [1]. The global installed capacity of renewable energy reached 2537 GW at the end of 2019, with hydropower leading the way (1190 GW globally), followed by wind energy (623 GW), and solar (586 GW) [5]. Explaining the simultaneous growth in demand for fossil fuels and growth in renewable energy adoption is a daunting cost differential; the overhead of capital required to begin renewable energy projects has been cited extensively as a barrier to renewable energy adoption, though the cost of clean energy relative to traditional energy has diminished substantially in recent years. Now, the overestimation of risk by financial institutions, the lack of legal and regulatory frameworks, and limitations on existing infrastructure and skills hinder the deployment of clean energy [6]. These obstacles apply particularly to nuclear energy [7].

Global energy market issues were intensified by the Russian invasion of Ukraine, as Russia is a top oil and gas exporter, and faced international sanctions, disrupting supply chains and threatening an energy crisis [8]. To achieve the deep decarbonization required to address these issues on a planetary scale, it is necessary to integrate low-carbon energy solutions into every sector of the economy while addressing the unprecedented need for speed and scaling, while balancing security, affordability, and sustainability [9].

The most recent research about creating carbon-neutral cities focuses on the potential roles of solar and wind energy in [10,11]. Some governments have been funding solar and wind farms [12,13] and financially penalizing emissions while giving tax rebates or loans to building owners to convert their current energy systems to be less carbon-intensive [14,15].

Nuclear reactors are operating in 32 countries and regenerate about 10% of the world's electricity, and fast-urbanizing countries such as India, the United Arab Emirates, and Pakistan have been pushing ahead with the construction of nuclear plants. In percentage, 9 of the top 10 countries generating electricity via nuclear reactors are in Europe, from France (62.6%) to Bulgaria (32.6%). Therefore, some of the most energy-intensive megacities or large metropolitan areas already depend, to varying degrees, on conventional nuclear power generated by large reactors distant sometimes hundreds of kilometers away—the list includes Rio de Janeiro, New York, London, Paris, Moscow, Delhi, Tokyo and Beijing.

Still, nuclear energy has been largely ignored as a source of clean energy in urban studies [16]. In the past few decades, partially due to the nuclear accidents of Chernobyl (1986) and Fukushima Daiichi (2011), several countries have shut down or paused their nuclear energy programs [17,18], and some scholars studying the energy transition have even welcomed the rollback of investment in nuclear.

However, nuclear energy has recently regained attention as the stakes of increasing carbon-free energy capacity become increasingly dire [19]. Partially responsible for the new wave of interest in nuclear energy technologies is the growing popularity of Small Modular Reactors (SMRs), which package nuclear reactors in smaller and simpler devices, thereby potentially lowering costs and increasing application flexibility [20,21].

This paper explores applications for this new wave of nuclear power systems by presenting use cases for one type of SMR, called Nuclear Batteries (NBs), which can autonomously generate between 1 and 20 MW of carbon-free heat, electricity, or both. We present a framework for the evaluation of use cases for NB deployment, focusing on social and environmental concerns, and on factors that make NB technology uniquely suited for some applications, specifically for large construction sites, informal settlements, grid resilience, emergency refuges, industrial power, and remote communities—cases which are often disconnected from the regular energy grid, or require substantial investments to be connected to the grid, or need large amounts of energy for short periods.

2. Nuclear Battery Technology

The renewed interest in SMR technology [22,23], has also boosted attention toward the class of systems known as Nuclear Batteries, but the phenomenon of a small nuclear reactor is not an unprecedented idea: SMRs (including small NBs) emerged in the early days of nuclear power [19]. The US Army experimented with terrestrial NBs in the 1960s, but the early devices were abandoned due to the high cost of development compared to the reliability of diesel generators.

Advances in nuclear technology, manufacturing methods, and supply lines as well as forthcoming opportunities for urban integration have made the adoption of SMRs, including NBs, more viable in recent years [20]. While nuclear energy has historically relied on the field construction of Gigawatt-scale reactors that can take decades and billions of dollars to complete, NBs represent a relatively low-cost alternative that can deliver nuclear energy through a modular and distributed infrastructure at a Megawatt scale. Some NBs can fit into a 6-m intermodal shipping container (ISO), be factory manufactured and fueled, and transported to use sites via existing shipping routes, becoming operational within days or weeks of arrival [24]. These devices operate semi-autonomously for five to ten years

before fuel resupply is necessary, after which they can be switched out for fresh devices and returned to the fueling facility for refurbishment [21].

NBs consist of a micro-reactor and turbine housed inside steel containment vessels that can fit within shipping containers. They meet all three requirements for nuclear reactor safety without operator intervention: (1) Rapid shutdown of the fission chain at the onset of anomalous conditions, (2) adequate cooling of the nuclear fuel during shutdown, (3) no uncontrolled release of nuclear material into the biosphere), and are refueled at centralized factory locations, eliminating the need for nuclear waste handling, long-term storage, or disposal at their operating sites. The spent nuclear fuel can be recycled or securely stored underground [25]. In this paper, we assume that each NB can produce 15 MW of heat and 5 MW of electricity consistently with refueling intervals of about 5–10 years.

Currently, there are no commercial nuclear batteries in operation. The leading designs, power, site and timeline of implementation, are MARVEL: 20 kWe, INL site, 2024; BWXT: 5 MWe, INL site, 2025; eVinci (Westinghouse): 5 MWe, INL site, 2026; and Hermes (Kairos): 35 MWt, Oak Ridge, 2026.

3. Nuclear Battery Evaluation Framework Methodology

We propose an evaluation framework to identify and evaluate characteristics of suitable use cases for NB technology. First, we describe the framework approach. Second, we explain how to apply the framework to energy systems. Lastly, we describe how this framework is situated within the energy literature.

3.1. Framework Background and Considerations

We began with a landscape analysis of needs for power sources in the tens of megawatts range for use cases that rely heavily on diesel generator configurations. Conceptual, qualitative, and quantitative factors are considered in the framework that evaluates the feasibility of a given NB use case, following a multi-criteria decision-making (MCDM) approach [26].

While the environmental impacts of renewable energy technologies are less harsh and far-reaching than fossil fuels, even clean energy technologies may negatively impact the environment and biodiversity [27]. To better understand the site-specific suitability of NBs compared to a different type of carbon-free electricity, the evaluation framework suggests taking into account local environmental effects [28,29]. We do not take into account the full life-cycle impacts of renewable energy technologies, as the framework is intended to serve as an aid in use case selection rather than perform a full environmental impact analysis of different options. Nevertheless, we acknowledge that this is a critical point to be taken into consideration in engineering and policy studies. The suitability of NBs for use in urban environments is considered through the "Site Attributes" factor.

The selection of factors included in the evaluation framework, shown in Table 1, was based on the wide range of factors present in the energy transition literature, including socio-technical system considerations [30], ability to escape from carbon lock-in in existing energy infrastructure [31], environmental policy [32], global political economy considerations [33], as well as on the system architecture of established energy evaluation frameworks such as the World Economic Forum's Energy Transition Index [34] and energy selection models for the United Nations Sustainable Development Goals (SDGs) [28]. Literature comparing different energy sources to each other for evaluation on multiple criteria was further reviewed to establish consistent definitions for each factor taken into account [29]. The twelve factors listed in the framework do not follow any hierarchy, rather they are factors that need to be satisfactorily addressed when assessing the feasibility of implementing nuclear batteries and may serve as guidelines in the decision-making process.

Table 1. Evaluation Factors and Units.

Factor	Unit	Definition
Installation Urgency	Days/Weeks/Months	Installation urgency defines the time sensitivity of the deployment, specifying the time frame in which an NB would need to arrive at the site.
Deployment Length	Weeks/Month/Years/Permanent	Deployment length refers to the amount of time that the NB would need to be operational at the site to provide a steady stream of energy.
Grid Connection	Yes/No	Grid connection states whether the NB would be connected to an existing local network, or if it would be a standalone source of power.
Site Attributes	User Input	Site attributes refer to a set of characteristics associated with the proposed location for the NB, such as whether it is space-constrained, remote, waterborne, faces extreme temperatures regularly, or challenges transporting the NB to the site.
MW/Site	Megawatt	MW/site estimates the energy demand (heat and/or electricity) of the use case based on a literature review of similar cases.
Global Site Count	10, 100, 1000, 1000+	Global site count is the estimated number of sites where the outlined use case for NBs could be applicable under scaled conditions.
Energy Market	(MW/Site multiplied by Global Site Count)	Energy market metric estimates how many global MW of power NBs could fill under the proposed use cases at scale.
People Served	100, 1000, 10,000, 100,000, 1,000,000, 1,000,000+	People served estimates the number of people NBs could provide heat and/or electricity to, based on population and industry data relevant to use case locations from public sources and scientific literature. Defined as 100+, 1000+, etc.
Social Barriers	User Input	Social barriers include security concerns, social opposition, and justice and equity concerns
Ecological Concerns	User Input	Ecological concerns center around possible disruption to land, water, or wildlife, and biodiversity.
Financial Constraints	User Input	Financial constraints look into funding for deployment, scalability economics, financial competitiveness of alternatives, and financial benefits to the community.
Most Competitive Alternatives	Diesel/Wind/Solar/Other	Most competitive alternatives are the status quo in terms of the current dominant energy technology for a given use case, based on a literature review.

3.2. Framework Architecture

Twelve factors were taken into account in the framework:

The framework's conceptual architecture, depicted in Figure 1, categorizes relevant factors, which can be combined and analyzed to generate a feasibility score. The framework's effectiveness was assessed through the evaluation of various use cases. Initially, three members of the research team proposed 10–12 use cases each, along with relevant information such as the use case context, real-world example, MW/site, people served per site, number of global sites anticipated, and the global energy market. Use cases that required MW output closest to the output of 1–3 NBs were given priority, as were cases that served a larger population and had the potential for scaling across many sites. The use cases were then reviewed by the group to identify any similarities and eliminate cases with insufficient information or those that were deemed too experimental.

A shortlist of cases was systematically evaluated using the standardized framework. Priority was given to cases where NBs would replace polluting energy sources such as diesel generators. Population data from academic literature and official records were used to determine the scale of people served. Social and environmental concerns were gathered from relevant academic literature and local government websites. Logistics considerations such as installation urgency and deployment duration were also taken into account, including the feasibility of nuclear battery refueling and delivery in the event of a natural disaster. The evaluation process aimed to identify the most viable energy alternative for each use case.

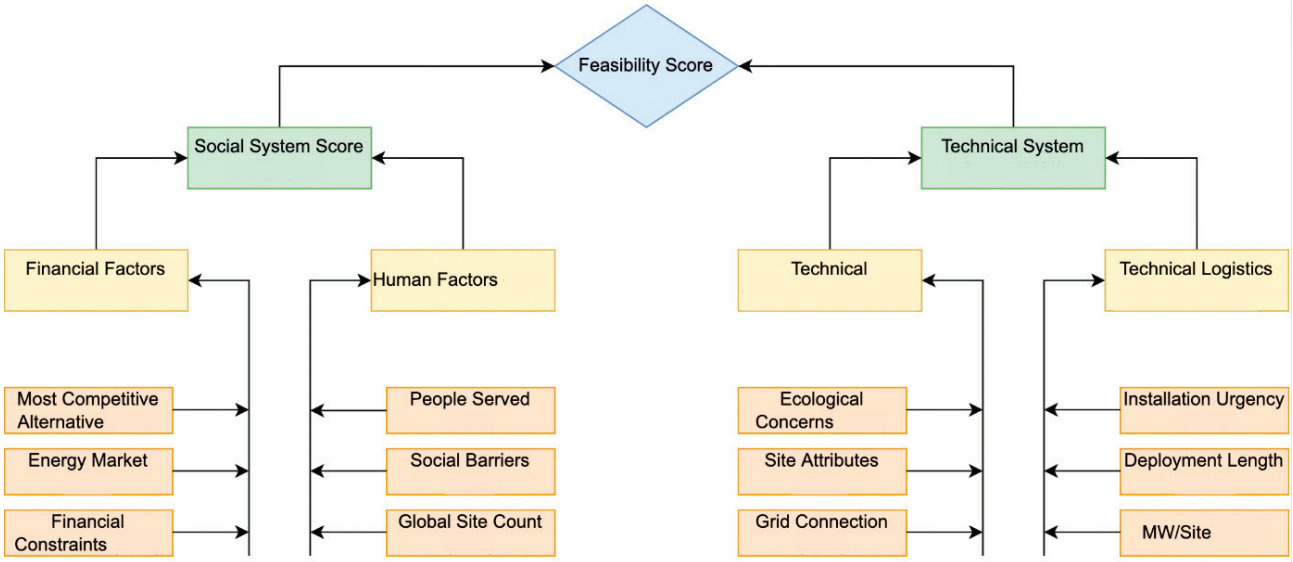


Figure 1. Evaluation Framework.

4. Use Cases

The following use cases focused on but not limited to the urban context, illustrate the breadth and depth of the potential impact of NBs. Each use case is technically feasible within the bounds of today's technology. The designs in this paper make the most of the technical qualities of these devices as clean, long-lasting, and energy-dense, as well as their formal characteristics as portable, replaceable, and small. The use cases locate NB-produced heat and electricity across sites that represent a spectrum of scales and contexts, but primarily center around urban or settlement-focused situations that are permanently or temporarily disconnected from the grid, or cases where temporary events would stress the grid. We also include use cases outside urban environments to provide insight into the applicability of this technology to different environments.

4.1. Construction

The construction industry takes up ~10% of the US GDP yet consumes only 1% of the electricity supply [35]. Out of 11 billion dollars spent on gas annually [36], USD 8 billion is spent on fuel to power mobile vehicles, while the other USD 3 billion is spent on static on-site equipment: concrete mixing stations, welding facilities, lighting, crane operations, and employee facilities. On average, 25% of the on-site energy is supplied through connections to the local grid, which burdens the domestic energy supply. The other 75% of the local energy supply is from portable diesel generators. The total energy consumption for various tools ranges from 7 kW for a single welding station or similar human-scale equipment to 244 kW for a large crane [37] and other heavy machinery, amounting to 10,400 kWh per year of operation.

Diesel generators are commonly exchanged between different construction sites, circulating in a closed market. In Canada, there are over 620,000 operable portable generators. Most of these construction diesel generators range from 1 to 225 kW [38]. The power costs

of nuclear batteries can be on par with traditional diesel generators if installation costs and deployment length are optimized [36]. While the energy consumption rates per generator and relative size of typical diesel generators are comparable to nuclear batteries, the cost efficiency of the latter improves with longer deployment length and consumption scale. Despite challenges posed by the shorter deployment of most construction sites, NBs could become competitive on large construction sites, pooling equipment sources, as well as maintaining a resilient fleet of exchange of nuclear batteries between construction sites at various stages of development.

Besides the on-site construction equipment, mobile vehicles directly involved in on-site activities can be efficiently powered with nuclear energy, provided they rely on electric charging. Because construction vehicles do not typically leave the boundaries of the construction site, special fueling services are arranged for individual construction objects. Such fueling stations require a regular supply of diesel in tanks, which poses separate logistical and environmental concerns. EV recharging facilities can be additionally powered by nuclear batteries which can further increase the utilization factor of nuclear energy on-site. Considering the average for all types of heavy machinery equipment, it is estimated that the TCO (Total cost of ownership) for electric-powered machinery will be on par with the diesel alternatives around 2025 [39], which makes a strong case for nuclear electricity generators as charging stations.

The construction use case responds mainly to the following factors within our framework: deployment length (ranging in months, large construction sites consume a substantial amount of energy during a relatively short period), grid connection (it does not need to be connected to the surrounding grid), site attributes (especially for large construction sites distant from consolidated urban areas), MW/Site (NB can supply the required high energy usage for a relatively short period without stressing the grid), energy market (NB can be moved to other construction sites easily), and ecological concerns (often large construction sites rely on diesel, especially for the machinery, to avoid stressing the local grid).

4.2. Informal Settlements

Informal urban settlements, also derogatorily referred to as "slums", persist as cities in the Global South grow rapidly. In total, 23.9% of the world's population lives in informal settlements [40] and 759 million people did not have access to electricity in 2019. Thus, any solution needs to consider two of the factors in our framework: global site count, in the order of thousands, and people served, in the order of millions.

Due to the political and regulatory hurdles complicating the delivery of electricity grid to informal settlements, energy access is most successfully facilitated by a wide range of medium-scale local stakeholders, from resident groups, and cartels [41] to community groups and NGOs [42], while top-down options to connect to the expanded grid prove generally unaffordable, challenging to organize, and face the challenges of damaged or stolen equipment or inequitable access and demand [43].

Even in cases in which solar-based microgrids are a viable option, investors may lack the motivation to install them due to the uncertain prospect of getting access to the central grid at any time, as well as dealing with complex maintenance, security, and vandalism [42]. Here, we have other factors in our framework: site attributes, which might hinder traditional energy-supply solutions, grid connection and, while a permanent solution is not achieved, deployment length. The possibility to install nuclear batteries temporarily (3–10 years [44]) and relocate them later makes it possible to use them as a temporary "infill" solution for electricity-deprived areas while the national grid takes time to adapt to the rising demand, not burdened by high operating and equipment security costs and the prospect of potentially unused infrastructure. Moreover, Nuclear Batteries are better adapted to handle variable energy loads and energy storage issues. A security fence might be necessary depending on the site's location.

With a population estimate of 200,000 to 700,000 and an area of two square kilometers [43], Kibera is one of Africa's largest informal settlements. Despite relative proximity

to downtown Nairobi's electricity grid, the upfront costs associated with connecting to the grid are unaffordable for the majority of residents. Alternative informal electricity suppliers provide undercover connections that are unstable, unsafe, and lead to major power disruptions in the main grid. Thus, charcoal and kerosene [40,43] are the main sources of fuel in Kibera, which are purchased and delivered in a decentralized manner—thus NBs should be assessed with competitive alternatives, which currently raise ecological and public health concerns.

Following the Nairobi Metro 2030 vision, the city is expected to expand its electricity infrastructure and energy grid [40]. With the influx of industry and population to the metropolitan area, the existing grid, already strained, faces increasing pressure. The Kenya Power and Lighting Company, a major actor in Nairobi's infrastructure space since 2007 [43], has been piloting subsidized connectivity programs and capping consumption per household at 40 kWh/month [45]. Community wishes and concerns about energy technologies [46] must be taken into account when applying the evaluation framework, as these use cases need community buy-in and enthusiasm to be realized.

4.3. Grid Resilience

The current intermittency issues with grid integration may hinder the large-scale adoption of renewable energy technologies [47]. Grid firming is a practice that addresses this by balancing the electric grid with other forms of stable energy to fill in the gaps [48], a practice relevant to both large-scale and microgrids.

The increasing frequency and intensity of natural disasters place additional stress on energy systems, affecting both energy demand and system resilience [49]. These consequences could be economically and socially costly for areas with high population densities such as cities [50]. The annual impact of weather-related blackouts in the United States has been estimated to be between USD 20 billion and USD 55 billion annually [51], although a single recent incident, the Texas cold snap, and resulting blackout, was estimated to cost USD 15–27 billion in direct economic costs alone [52]. Reliable sources of energy can power critical infrastructure in emergencies through a variety of strategies, one of which could be small modular reactors or even NBs.

Using nuclear batteries to support grid-firming is considerate of large-scale power outages, but NBs could also be used to provide power for neighborhoods that are affected by weather-related blackouts. When not used for emergency power, the NBs would need to be used for other applications to ensure a good lifecycle utilization of these assets. This use case responds to factors such as grid connection, MW/site (or variation in energy demand), and financial constraints (increasing energy generation at this scale requires work that might take several years with large spatial footprints).

The site considered for this application is an existing power generation facility, so social concerns about proximity to the nuclear battery are lower. Ecological concerns are minimal, as the NBs will be sited in an area already being used for energy generation. Further, nuclear batteries are expected to be cost-competitive with industrial diesel generators, when used with a high-capacity factor, and have the added benefit of not contributing to greenhouse gas emissions or poor air quality.

During the Texas 2021 cold snap, 26,000 MW of electricity load was wiped off the grid [52], with millions of people losing power and over 240 dying as a direct or indirect result [53,54]. This incident presented an unusually dire energy situation as the overall problem was not downed power lines or other issues common in storms, but rather that every power generation source contributing to the energy mix on Texas' grid was underperforming in the cold due to poor winterization [55]. Like other energy sources that Texas was reliant upon—primarily natural gas, but also including coal and wind—nuclear also underperformed during the storm, but this was partially due to one of four nuclear reactors going offline due to a feedwater pump issue. Nuclear batteries do not rely on water as a coolant or working fluid in their operation [55].

4.4. Emergency Refuge

As natural disasters increase in frequency and intensity due to climate change, there is an increased risk of population displacement [56]. In total, 916,000 people were displaced by disasters in the US alone in 2019. Sports stadiums, conference centers, and other large buildings have been used to provide shelter to populations immediately impacted by these events. These structures are often located in downtown areas, and emergency shelters need to be easily accessible [57]. Stadiums and convention centers will also typically have large parking lots, which could be good sites for nuclear battery installation—and respond directly to the factor's installation urgency (it must happen in matters of hours, of days) and deployment length (NBs are needed during the critical phase, before the grid is rebuilt).

In the wake of natural disasters, the power source for a shelter facility needs to be robust as it faces environmental security challenges, including but not limited to emergencies such as ongoing earthquake aftershocks, flooding, fires, deep snow, extreme heat, or other dangerous weather conditions. The power demands vary, but sports stadiums on game days may use up to 10 MW. Convention centers use a lower amount of power per day but are continuously used. The Pennsylvania Convention Center in Philadelphia reported using roughly 24,000,000 kWh of electricity over the course of 2019, which would equate to roughly 2.7 MWh of continuous operation.

Stadiums, conference centers, and similar facilities tend to be in commercial districts surrounded by large parking lots or in open spaces outside the city, so the social concerns may not be as severe as if within dense residential areas. Ecological concerns are minimal since nuclear batteries are emission-free with a container-size footprint and are equipped with radiation shielding. It is possible that for both general and environmental security reasons, such as protection from debris dislodged by strong winds, flooding, and more, these nuclear batteries could be embedded below grade.

Financial constraints associated with this use case could be mitigated by federal, state, or private funding. For instance, in the United States, the Disaster Relief Fund (DRF) is a congressional budget overseen by the Federal Emergency Management Agency (FEMA), but it is not designed for resilience, only response [58]. FEMA also has grants for disaster preparedness but seems more driven toward research than toward infrastructure installation [56].

In the wake of Hurricane Katrina's impact on the United States Gulf Coast in, 2005, the New Orleans Superdome became a refuge of last resort. During the hurricane, the Superdome was running on its backup generator within one day of the storm making landfall, over the course of 5 days, 27–31 August 2005. While thousands of people were saved by taking refuge inside it, conditions were awful—there was no running water, not enough light, and no air conditioning—partially because of energy limitations as the facility was operating on its backup generator. This generator also had to be refueled with oil during the emergency, which was especially challenging to access due to floodwaters. In this instance, nuclear batteries which can be deployed in days, possibly less than a day if the system is delivered fully functional and pre-assembled on a barge or container, would have been able to provide more energy without refueling, thereby improving the conditions and stability of an emergency substantially.

4.5. Industrial Power

Roughly a quarter of global greenhouse gas emissions are induced by the industrial sector [59], which is primarily powered by coal and natural gas. The interdependence between power-hungry industrial facilities and the fossil fuel industry represents a serious challenge to overcome for the desired energy transition [60]. It also complicates the logistics of factory constructions that have to be situated closer to large-scale, expensive power plants or initiate the long-term construction of new energy grids and power sources. This geographic co-dependence is further aggravated in remote or sparsely populated areas with few prominent industries, such as aluminum production and hydropower in Iceland [61]—while aluminum can be produced with carbon-free electricity, the scalability of the process

is limited by geographical constraints. As heavy industries are tied to major power sources and not to major nodes of supply or transportation, the additional logistical expenses further aggravate their carbon impact.

In steel making, a key part of the process is to use a coal-fired blast furnace to strip oxygen from iron ore, creating a product called pig iron that is then turned into steel in an electric arc furnace. The blast furnace stage can be replaced by an electrolysis process, similar to the process used to make aluminum. There are early attempts to decarbonize steel production with other renewable energy sources that face challenges of scale and deployment. [62] Nucor's Sedalia micro-mill, the world's first wind-powered steel factory, consumes 55 MW of power and has required the establishment of a USD 250 million wind farm in Kansas that transfers electricity to Nucor's mill in Missouri [63]. The costs of coordinating and co-locating two infrastructure projects, the uncertainty of wind power output, and the construction schedule can be alleviated with only six 10-MW nuclear batteries capable of powering a similar-scale plant for several years, within several days from deployment. The constant high-volume demand for energy characteristic of heavy industrial plants can take full advantage of energy-intensive and compact nuclear and thus tackle the initial high installation costs.

In addition to decarbonizing conventional carbon-heavy industries, nuclear batteries can effectively power novel "green heavy industries". One of the prospective applications is the joint installation of nuclear batteries with hydrogen facilities. Like nuclear batteries, hydrogen power plants are actively investigated as novel energy sources for heavy industries [64]. However, the operation of a hydrogen plant itself necessitates a significant energy input which brings installation challenges similar to those of conventional plants, but progress is being made—in Denmark, a green hydrogen testing plant is being powered by a 3 MW wind turbine [65]. The compactness and speed of installation of nuclear batteries can allow the deployment of such hydrogen production facilities far beyond the regions with abundant renewable energy supplies.

Nuclear batteries' introduction to the heavy industry sector may be particularly promising due to lower pressure from social and environmental concerns. The air quality and other health impacts of heavy industrial facilities, and their high land and infrastructure footprint significantly outnumber the technical challenges associated with nuclear microreactors and may be reduced with the introduction of compact, reliable, carbon-free energy sources like nuclear batteries.

4.6. Remote Communities

Remote communities often face barriers to electricity access. In 2020, ~800 million people remained without access to electricity [66], with roughly 84% residing in rural areas [67]. Remote communities typically have microgrid power systems ranging from 200 kW to 5 megawatts (MW). "Last-mile electrification" is an exceedingly complex challenge for electricity providers to overcome, spanning regulatory, policy, and technological aspects. Remote and rural communities worldwide often rely on diesel generators or other polluting sources to generate electricity and heat [68]. A business-as-usual approach is unlikely to improve electrification rates significantly.

Electricity access has traditionally been pursued through grid expansion programs [69], which may not account for unique factors and limitations present in rural or remote communities [67]. Alternate technological frameworks such as mini- or micro-grids, solar home systems, or diesel generators may better fit into the needs of a remote community. However, communities primarily powered by diesel generators are susceptible to supply chain disruptions for their primary power source, as well as fuel price volatility [70]. Additionally, there is an urgent need to reduce carbon emissions in grids that rely on coal, oil, and natural gas. This need, coupled with the rapidly decreasing costs of renewable and carbon-free electricity, the necessity of high reliability in remote locations, and the difficulty of expanding centralized grid infrastructure to remote locations favor a distributed generation approach [71].

Rural communities in Canada, home to over 6.5 million people [72] are not situated close enough to each other or other large towns to warrant a connected electrical grid. In Canada, 70% of remote communities rely on diesel generators to produce electricity, 13% on hydro, and 17% use a combination of other fossil fuels [73]. In Nunavut, which forms much of the Canadian Arctic Archipelago, there are ~26 remote communities, representing a total population of around 20,000 people [72]. All of Nunavut's off-grid communities run on diesel power, requiring about 55,000,000 L of diesel per year for an electrical capacity of about 78 MW in total [74]. Nunavut consumes the least total energy of any of Canada's provinces, and its per capita energy demands are also the smallest [74], but it is routinely exposed to the risks of fuel transport: over the last five years, Nunavut has experienced at least 549 fuel spills, averaging a spill almost every three days [75].

Replacing diesel with NBs in Nunavut's fuel infrastructure could deliver major benefits at a minor conversion cost: the existing trucks and vessels that transport Nunavut's 55 million liters of fuel each year could be retrofitted to carry shipping-container-sized NBs, and bulk fuel facilities, or "tank farms", could be converted to NB storage or exchange centers. Almost every town in Nunavut could be sustained with only a 5 MW NB; the only exception is the territory's capital Iqaluit, which could be served by a set of NBs adding up to 20 MW [76]. This change would minimize the risk of fuel spills and diminish the overall stress on Nunavut's endangered Arctic ecoregion by replacing a months-long season of intensive fuel transport every year with NB exchange operations every 5 to 10 years. A reduced number of diesel generators may need to be retained as a contingency plan in the unlikely event of NB failure.

5. Conclusions

Nuclear Batteries offer a potential solution to the limitations posed by the politics of energy investment and regulation on traditional nuclear power plants, enabling more efficient and manageable modes of construction, distribution, and management. However, significant regulatory uncertainty remains about several aspects of NBs and their deployment. The scarcity of operational data presents a significant barrier to regulatory assessments in the area, especially given the vast number of environments in which NBs can function, including urban environments, where special licensing considerations may be needed, and remote areas, where NBs will largely be operating autonomously. Another key consideration is shipment, including whether to ship NBs with fuel or transport them separately, which requires insight into transportation limitations and requirements, as well as the implementation of processes around delivery, audit, and storage. Answering these questions is critical to suitably addressing considerations around refueling, a key aspect of NBs' unique value proposition as an energy source.

Considerations also remain for the upfront costs of NB fabrication and deployment, and pathways to reduce cost through factory fabrication to allow the technology to scale, which further merits exploration of business models for commercialization in domestic and international markets [77].

Lastly, ensuring the security of NB energy sources and any necessary safeguards for the technology is crucial. While physical safety is an essential component of NB design (including a robust fuel-reactor package, the possibility of remote monitoring, and the lack of possibility of runaway reactions), cybersecurity measures and attention to physical damage to the NB containers must be implemented [77].

The use cases presented in this paper demonstrate the persuasive ability of NBs not only to wedge themselves into existing energy uses but also to drive strategies for multimodal environmentally sustainable infrastructure in the urban sphere. The use cases discussed here were selected based on a twelve-factor framework we propose as a guideline to assess situations in which the implementation of nuclear batteries presents clear advantages over currently used alternatives. Unlike traditional nuclear power plants, which demand the configuration of cumbersome policies and financial structures around their needs, NBs embody a new paradigm of energy adaptability, potentially not only lightening

the political burden of clean energy integration but also creating new opportunities for the emergence of unconventional hybrid development between infrastructure and industry. We illustrate their relevance to climate change mitigation and adaptation, in the form of designs for their use in humanitarian relief and remote settlements. Finally, we reveal the advantages of NBs over those technologies in terms of energy density and relative ecological unobtrusiveness, being an alternative to outdated systems whose drawbacks have historically prevented nuclear energy from making its full potential contribution to global decarbonization.

Author Contributions: Conceptualization, F.D. and J.B.; methodology, S.P., M.K., F.D., C.G. and J.B.; investigation, S.P., M.K., F.D., C.C. and C.G.; writing—original draft preparation, S.P., M.K., C.C. and C.G.; writing—review and editing, S.P., M.K., F.D., C.C., C.G. and J.B.; supervision, F.D., J.B. and C.R.; project administration, F.D.; funding acquisition, J.B., F.D. and C.R. All authors have read and agreed to the published version of the manuscript.

Funding: Funding for this project was obtained from the BT Charitable Foundation. The authors would also like to thank the Dubai Future Foundation, Toyota, UnipolTech, Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria, Tele2, Volkswagen Group America, FAE Technology, Fédération Internationale de l'Automobile, R io de Janeiro, Emilia-Romagna, Stockholm, Amsterdam, and all members of MIT Senseable City Lab Consortium.

Data Availability Statement: No new data has been generated during this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. EIA. *EIA Projects Nearly 50% Increase in World Energy Usage by 2050, Led by Growth in Asia*; U.S. Energy Information Administration: Washington, DC, USA, 2019. Available online: https://www.eia.gov/todayinenergy/detail.php?id=41433 (accessed on 1 December 2023).
2. Intergovernmental Panel on Climate Change. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate, Change*; Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., et al., Eds.; Cambridge University Press: Cambridge, UK; New York, NY, USA, *in press*. Available online: https://report.ipcc.ch/ar6wg1/pdf/IPCC_AR6_WG1_FinalDraft_FullReport.pdf (accessed on 1 December 2023). [CrossRef]
3. International Energy Agency (IEA). *Coal 2022: Analysis and Forecast to 2025*; International Energy Agency: Washington, DC, USA, 2022. Available online: https://iea.blob.core.windows.net/assets/91982b4e-26dc-41d5-88b1-4c47ea436882/Coal2022.pdf (accessed on 1 December 2023).
4. IEA. *Net Zero by 2050—Analysis*; International Energy Agency: Paris, France, 2021. Available online: https://www.iea.org/reports/net-zero-by-2050 (accessed on 24 November 2023).
5. IRENA. *Renewable Capacity Highlights*; International Renewable Energy Agency: Abu Dhabi, United Arab Emirates, 2020. Available online: https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2020/Mar/IRENA_RE_Capacity_Highlights_2020.pdf?la=en&hash=B6BDF8C3306D271327729B9F9C9AF5F1274FE30B (accessed on 1 December 2023).
6. Safwat Kabel, T.; Bassim, M.; Reasons for Shifting and Barriers to Renewable Energy: A Literature Review. SSRN Scholarly Paper (ID 3535783), Social Science Research Network (2020). Available online: https://papers.ssrn.com/abstract=3535783 (accessed on 1 December 2023).
7. Handrlica, J. Facilitating Deployment of Transportable Nuclear Power Plants Through a New Regime of Mutual Recognition. *J. World Energy Law Bus.* **2022**, *15*, 282–294. [CrossRef]
8. Ozili, P.K.; Global Economic Consequence of Russian Invasion of Ukraine. Social Science Research Network. 2022. Available online: https://ssrn.com/abstract=4064770 (accessed on 1 December 2023).
9. Heffron, R.J.; McCauley, D.; de Rubens, G.Z. Balancing the Energy Trilemma Through the Energy Justice Metric. *Appl. Energy* **2018**, *229*, 1191–1201. [CrossRef]
10. Karvonen, A.; Guy, S. Urban Energy Landscapes and the Rise of Heat Networks in the United Kingdom. *J. Urban Technol.* **2018**, *25*, 19–38. [CrossRef]
11. Stewart, I.D.; Kennedy, C.A.; Facchini, A.; Mele, R. The Electric City as a Solution to Sustainable Urban Development. *J. Urban Technol.* **2018**, *25*, 3–20. [CrossRef]
12. Solar Energy Technologies Office. *Homeowner's Guide to the Federal Tax Credit for Solar Photovoltaics*; Energy.Gov, Office of Energy Efficiency & Renewable Energy: Washington, DC, USA, 2020. Available online: https://www.energy.gov/eere/solar/homeowners-guide-federal-tax-credit-solar-photovoltaics (accessed on 1 December 2023).

13. Wind Energy Technologies Office. *U.S. Wind Industry Federal Incentives, Funding, and Partnership Opportunities Fact Sheet*; Energy.Gov, Office of Energy Efficiency & Renewable Energy: Washington, DC, USA, 2021. Available online: https://www.energy.gov/eere/wind/articles/us-wind-industry-federal-incentives-funding-and-partnership-opportunities-fact (accessed on 1 December 2023).
14. Zhang, F.; Chung, C.K.L.; Lu, T.; Wu, F. The Role of the Local Government in China's Urban Sustainability Transition: A Case Study of Wuxi's Solar Development. *Cities* **2021**, *117*, 103294. [CrossRef]
15. Lee, J.-S.; Kim, J.-W. South Korea's Urban Green Energy Strategies: Policy Framework and Local Responses under the Green Growth. *Cities* **2016**, *54*, 20–27. [CrossRef]
16. Rutherford, J.; Coutard, O. Urban Energy Transitions: Places, Processes and Politics of Socio-Technical Change. *Urban Stud.* **2014**, *51*, 1353–1377. [CrossRef]
17. Aoyagi, M. The Impact of the Fukushima Accident on Nuclear Power Policy in Japan. *Nat. Energy* **2021**, *6*, 326–328. [CrossRef]
18. Rinscheid, A.; Wüstenhagen, R. Germany's Decision to Phase out Coal by 2038 Lags behind Citizens' Timing Preferences. *Nat. Energy* **2019**, *4*, 856–863. [CrossRef]
19. Ramana, M.V. The Forgotten History of Small Nuclear Reactors. *IEEE Spectr.* **2015**, *27*, 44–58. Available online: https://spectrum.ieee.org/the-forgotten-history-of-small-nuclear-reactors (accessed on 24 November 2023). [CrossRef]
20. Black, G.; Shropshire, D.; Araújo, K.; van Heek, A. Prospects for Nuclear Microreactors: A Review of the Technology, Economics, and Regulatory Considerations. *Nuclear Technol.* **2023**, *209* (Suppl. 1), S1–S20. [CrossRef]
21. Vujić, J.; Bergmann, R.M.; Škoda, R.; Miletić, M. Small Modular Reactors: Simpler, Safer, Cheaper? *Energy* **2012**, *45*, 288–295. [CrossRef]
22. Liu, Y.; Huang, G.; Chen, J.; Zhang, X.; Zheng, X.; Zhai, M. Development of an Optimization-Aided Small Modular Reactor Siting Model—A Case Study of Saskatchewan, Canada. *Appl. Energy* **2022**, *305*, 117867. [CrossRef]
23. Nian, V.; Ghori, A.; Guerra, E.M.; Locatelli, G.; Murphy, P. Accelerating Safe Small Modular Reactor Development in Southeast Asia. *Util. Policy* **2022**, *74*, 101330. [CrossRef]
24. Black, G.; Black, M.A.T.; Solan, D.; Shropshire, D. Carbon Free Energy Development and the Role of Small Modular Reactors: A Review and Decision Framework for Deployment in Developing Countries. *Renew. Sustain. Energy Rev.* **2015**, *43*, 83–94. [CrossRef]
25. Lerner, L. *Nuclear Fuel Recycling Could Offer Plentiful Energy*; Argonne National Laboratory: Lemont, IL, USA, 2012. Available online: https://www.anl.gov/article/nuclear-fuel-recycling-could-offer-plentiful-energy (accessed on 24 November 2023).
26. Çelikbilek, Y.; Tüysüz, F. An Integrated Grey Based Multi-Criteria Decision Making Approach for the Evaluation of Renewable Energy Sources. *Energy* **2016**, *115*, 1246–1258. [CrossRef]
27. Sayed, E.T.; Wilberforce, T.; Elsaid, K.; Rabaia MK, H.; Abdelkareem, M.A.; Chae, K.J.; Olabi, A.G. A Critical Review on Environmental Impacts of Renewable Energy Systems and Mitigation Strategies: Wind, Hydro, Biomass, and Geothermal. *Sci. Total Environ.* **2021**, *766*, 144505. [CrossRef]
28. Büyüközkan, G.; Karabulut, Y.; Mukul, E. A Novel Renewable Energy Selection Model for United Nations' Sustainable Development Goals. *Energy* **2018**, *165*, 290–302. [CrossRef]
29. Kaldellis, J.K.; Apostolou, D.; Kapsali, M.; Kondili, E. Environmental and Social Footprint of Offshore Wind Energy: Comparison with Onshore Counterpart. *Renew. Energy* **2016**, *92*, 543–556. [CrossRef]
30. Li, F.G.N.; Trutnevyte, E.; Strachan, N. A Review of Socio-Technical Energy Transition (STET) Models. *Technol. Forecast. Soc. Chang.* **2015**, *100*, 290–305. [CrossRef]
31. Seto, K.C.; Davis, S.J.; Mitchell, R.B.; Stokes, E.C.; Unruh, G.; Ürge-Vorsatz, D. Carbon Lock-In: Types, Causes, and Policy Implications. *Annu. Rev. Environ. Resour.* **2016**, *41*, 425–452. [CrossRef]
32. Bashir, M.F.; Sadiq, M.; Talbi, B.; Shahzad, L. An Outlook on the Development of Renewable Energy, Policy Measures to Reshape the Current Energy Mix, and How to Achieve Sustainable Economic Growth in the Post COVID-19 Era. *Environ. Sci. Pollut. Res.* **2022**, *29*, 43636–43647. [CrossRef] [PubMed]
33. Pearse, R. Theorising the Political Economy of Energy Transformations: Agency, Structure, Space, Process. *New Political Econ.* **2021**, *26*, 951–963. [CrossRef]
34. Singh, H.V.; Bocca, R.; Gomez, P.; Dahlke, S.; Bazilian, M. The Energy Transitions Index: An Analytic Framework for Understanding the Evolving Global Energy System. *Energy Strategy Rev.* **2019**, *26*, 100382. [CrossRef]
35. Sharrard, A.; Matthews, H.S.; Roth, M. Environmental Implications of Construction Site Energy Use and Electricity Generation. *J. Constr. Eng. Manag.* **2007**, *133*, 846–854. [CrossRef]
36. Zhang, Y.; Yan, D.; Hu, S.; Guo, S. Modelling of Energy Consumption and Carbon Emission from the Building Construction Sector in China: A Process-Based LCA Approach. *Energy Policy* **2019**, *134*, 110949. [CrossRef]
37. Taher, A.H.; Elbeltagi, E.E. Integrating Building Information Modeling with Value Engineering to Facilitate the Selection of Building Design Alternatives Considering Sustainability. *Int. J. Constr. Manag.* **2021**, *23*, 1886–1901. [CrossRef]
38. Biswas, W.K. Carbon Footprint and Embodied Energy Consumption Assessment of Building Construction Works in Western Australia. *Int. J. Sustain. Built Environ.* **2014**, *3*, 179–186. [CrossRef]
39. McKinsey. *Electrifying Heavy Machinery and Equipment*; McKinsey: New York, NY, USA, 2019. Available online: https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/harnessing-momentum-for-electrification-in-heavy-machinery-and-equipment (accessed on 1 December 2023).

40. United Nations. Goal 11: Sustainable Cities and Communities. 2021. Available online: https://unstats.un.org/sdgs/report/2021/goal-11/ (accessed on 1 December 2023).
41. de Bercegol, R.; Monstadt, J. The Kenya Slum Electrification Program: Local Politics of Electricity Networks in Kibera. *Energy Res. Soc. Sci.* **2018**, *41*, 249–258. [CrossRef]
42. Pearce, F. In Rural India, Solar-Powered Microgrids Show Mixed Success. *Yale Environment 360*, 14 January 2016. Available online: https://e360.yale.edu/features/in_rural_india_solar-powered_microgrids_show_mixed_success (accessed on 1 December 2023).
43. Yonemitsu, A.; Njenga, M.; Iiyama, M.; Matsushita, S. Household Fuel Consumption Based on Multiple Fuel Use Strategies: A Case Study in Kibera Slums. *APCBEE Procedia* **2014**, *10*, 331–340. [CrossRef]
44. Testoni, R.; Bersano, A.; Segantin, S. Review of Nuclear Microreactors: Status, Potentialities and Challenges. *Prog. Nucl. Energy* **2021**, *138*, 103822. [CrossRef]
45. IRENA. *Electrification Strategies for Slum Customers, KENYA*; International Renewable Energy Agency: Naples, Italy, 2012. Available online: https://www.irena.org/-/media/Files/IRENA/Agency/Events/2012/Sep/2/Harun_Mwangi.pdf?la=en&hash=87CB59AFB25437F730EA6BF6F3CCC2D84F71F993 (accessed on 1 December 2023).
46. Gross, C. Community Perspectives of Wind Energy in Australia: The Application of a Justice and Community Fairness Framework to Increase Social Acceptance. *Energy Policy* **2007**, *35*, 2727–2736. [CrossRef]
47. Naeem, A.; Hassan, N.U. Renewable Energy Intermittency Mitigation in Microgrids: State-of-the-Art and Future Prospects. In Proceedings of the 4th International Conference on Green Energy and Applications (ICGEA), Singapore, 7–9 March 2020.
48. Quann, C.; Bradley, T.H. Renewables Firming Using Grid Scale Battery Storage in a Real-Time Pricing Market. In Proceedings of the IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, USA, 23–26 April 2017.
49. Perera, A.T.D.; Nik, V.M.; Chen, D.; Scartezzini, J.L.; Hong, T. Quantifying the Impacts of Climate Change and Extreme Climate Events on Energy Systems. *Nat. Energy* **2020**, *5*, 150–159. [CrossRef]
50. World Economic Forum. *The Global Risks Report 2016*; World Economic Forum: Cologny, Switzerland, 2016. Available online: https://www.weforum.org/reports/the-global-risks-report-2016/ (accessed on 1 December 2023).
51. Kenward, A.; Raja, U. *Blackout: Extreme Weather, Climate Change and Power Outages*; Climate Central: Princeton, NJ, USA, 2014. Available online: https://assets.climatecentral.org/pdfs/PowerOutages.pdf (accessed on 1 December 2023).
52. Dean, K.M. *Texas Cold Snap Highlights Need for Improved Power Systems*; Argonne National Laboratory: Lemont, IL, USA, 2022. Available online: https://www.anl.gov/article/texas-cold-snap-highlights-need-for-improved-power-systems (accessed on 1 December 2023).
53. Svitek, P. Texas Puts Final Estimate of Winter Storm Death Toll at 246. *The Texas Tribune*, 2 January 2022. Available online: https://www.texastribune.org/2022/01/02/texas-winter-storm-final-death-toll-246/ (accessed on 1 December 2023).
54. Doan, L. How Many Millions Are Without Power in Texas? It's Impossible to Know for Sure. *Time*, 17 February 2021. Available online: https://time.com/5940232/millions-without-power-texas/ (accessed on 1 December 2023).
55. Busby, J.W.; Baker, K.; Bazilian, M.D.; Gilbert, A.Q.; Grubert, E.; Rai, V.; Rhodes, J.D.; Shidore, S.; Smith, C.A.; Webber, M.E. Cascading Risks: Understanding the 2021 Winter Blackout in Texas. *Energy Res. Soc. Sci.* **2021**, *77*, 102106. [CrossRef]
56. Palinkas, L.A. *Global Climate Change, Population Displacement, and Public Health: The Next Wave of Migration*; Springer International Publishing: Cham, Switzerland, 2020. Available online: https://link.springer.com/book/10.1007/978-3-030-41890-8 (accessed on 1 December 2023).
57. IDMC. *United States of America Figure Analysis—Displacement Related to Disasters*; Internal Displacement Monitoring Centre: Geneva, Switzerland, 2019. Available online: https://www.internal-displacement.org/sites/default/files/inline-files/GRID-2019-Disasters-Figure-Analysis-UnitedStates.pdf (accessed on 1 December 2023).
58. FEMA. *Disaster Relief Fund: Monthly Reports*; FEMA: Washington, DC, USA, 2023. Available online: https://www.fema.gov/about/reports-and-data/disaster-relief-fund-monthly-reports (accessed on 1 December 2023).
59. EPA. *Greenhouse Gas Emissions*; United States Environmental Protection Agency: Washington, DC, USA, 2021. Available online: https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions#industry (accessed on 1 December 2023).
60. Michaelowa, A. The Politics of Climate Change in Germany: Ambition versus Lobby Power. *WIREs Clim. Chang.* **2013**, *4*, 315–320. [CrossRef]
61. Overall Aluminium Industry. Aluminium Industry in Iceland. *Alum. Int. Today* **2021**, *34*, 7–9. Available online: https://www.proquest.com/scholarly-journals/aluminium-industry-iceland/docview/2583135096/se-2 (accessed on 1 December 2023).
62. EU-JRC. *Decarbonisation of Industrial Heat: The Iron and Steel Sector*; The European Union Joint Research Centre: Brussels, Belgium, 2020. Available online: https://ee-ip.org/fileadmin/user_upload/DOCUMENTS/Content/JRC119415_-_Iron_and_steel_decarbonisation_brief.pdf (accessed on 24 November 2023).
63. Hinton, D. *Wind Power Is Helping Meet World's Need for Sustainable Steel*; World Steel Association: Brussels, Belgium, 2022. Available online: https://worldsteel.org/steel-stories/infrastructure/wind-power-meeting-worlds-need-for-sustainable-steel/ (accessed on 1 December 2023).
64. Hoffmann, C. *Decarbonization Challenge for Steel*; McKinsey: New York, NY, USA, 2020. Available online: https://www.mckinsey.com/~/media/McKinsey/Industries/Metals%20and%20Mining/Our%20Insights/Decarbonization%20challenge%20for%20steel/Decarbonization-challenge-for-steel.pdf (accessed on 1 December 2023).

65. Siemens. *Green Hydrogen*; Siemens Gamesa: Munich, Germany, 2023. Available online: https://www.siemensgamesa.com/products-and-services/hybrid-and-storage/green-hydrogen (accessed on 1 December 2023).
66. Garces, E. Lessons from Last Mile Electrification in Colombia: Examining the Policy Framework and Outcomes for Sustainability. *Energy Res. Soc. Sci.* **2021**, *79*, 102156. [CrossRef]
67. Azimoh, C.L.; Klintenberg, P.; Mbohwa, C.; Wallin, F. Replicability and Scalability of Mini-Grid Solution to Rural Electrification Programs in Sub-Saharan Africa. *Renew. Energy* **2017**, *106*, 222–223. [CrossRef]
68. Natural Resources Canada. *Clean Energy for Rural and Remote Communities Program*; Natural Resources Canada: Ottawa, ON, Canada, 2023. Available online: https://natural-resources.canada.ca/reducingdiesel (accessed on 1 December 2023).
69. Ilskog, E.; Kjellström, B. And Then They Lived Sustainably Ever After? Assessment of Rural Electrification Cases by Means of Indicators. *Energy Policy* **2008**, *36*, 2674–2684. [CrossRef]
70. Office of Energy Efficiency and Renewable Energy. *Community-Scale Isolated Power Systems*; U.S. Department of Energy: Washington, DC, USA, 2019. Available online: https://www.energy.gov/sites/default/files/2019/09/f66/73355-9.pdf (accessed on 1 December 2023).
71. IRENA. *Off-Grid Renewable Energy Systems: Status and Methodological Issues*; International Renewable Energy Agency: Abu Dhabi, United Arab Emirates, 2015. Available online: https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2015/IRENA_Off-grid_Renewable_Systems_WP_2015.pdf (accessed on 1 December 2023).
72. Statistics Canada. *Population Growth in Canada's Rural Areas, 2016 to 2021*; Statistics Canada: Ottawa, ON, Canada, 2021. Available online: https://www12.statcan.gc.ca/census-recensement/2021/as-sa/98-200-x/2021002/98-200-x2021002-eng.cfm (accessed on 1 December 2023).
73. The Atlas of Canada. Remote Communities Energy Database. Available online: https://atlas.gc.ca/rced-bdece/en/index.html (accessed on 1 December 2023).
74. CER. *Provincial and Territorial Energy Profiles—Nunavut*; CER—Canada Energy Regulator: Calgary, AB, Canada, 2018. Available online: https://www.cer-rec.gc.ca/en/data-analysis/energy-markets/provincial-territorial-energy-profiles/provincial-territorial-energy-profiles-nunavut.html (accessed on 24 November 2023).
75. Environment and Natural Resources, Government of Northwest Territories. Spills. Available online: https://www.enr.gov.nt.ca/en/spills (accessed on 1 December 2023).
76. Aboriginal Affairs and Northern Development Canada (AANDC), and Natural Resources Canada (NRCan). *Status of Remote/Off-Grid Communities in Canada*; Government of Canada: Ottawa, ON, USA, 2011. Available online: https://www.nrcan.gc.ca/sites/www.nrcan.gc.ca/files/canmetenergy/files/pubs/2013-118_en.pdf (accessed on 24 November 2023).
77. Buongiorno, J.; Carmichael, B.; Dunkin, B.; Parsons, J.; Smit, D. Can Nuclear Batteries Be Economically Competitive in Large Markets? *Energies* **2021**, *14*, 4385. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

urban science

MDPI

Article

Scatter Search Algorithm for a Waste Collection Problem in an Argentine Case Study

Diego Rossit [1,2,*], Begoña González Landín [3,*], Mariano Frutos [1,4] and Máximo Méndez Babey [3]

[1] Department of Engineering, Universidad Nacional del Sur, Bahía Blanca 8000, Argentina; mfrutos@uns.edu.ar
[2] Instituto de Matemática, Universidad Nacional del Sur (UNS)—CONICET, Bahía Blanca 8000, Argentina
[3] Instituto Universitario SIANI, Universidad de Las Palmas de Gran Canaria, 35017 Las Palmas de Gran Canaria, Spain; maximo.mendez@ulpgc.es
[4] Instituto de Investigaciones Económicas y Sociales del Sur, Universidad Nacional del Sur (UNS)—CONICET, Bahía Blanca 8000, Argentina
* Correspondence: diego.rossit@uns.edu.ar (D.R.); bego.landin@ulpgc.es (B.G.L.)

Abstract: Increasing urbanization and rising consumption rates are putting pressure on urban systems to efficiently manage Municipal Solid Waste (MSW). Waste collection, in particular, is one of the most challenging aspects of MSW management. Therefore, developing computer-aided tools to support decision-makers is crucial. In this paper, a Scatter Search algorithm is proposed to address the waste collection problem. The literature is relatively scarce in applying this algorithm, which has proven to be efficient in other routing problems, to real waste management problems. Results from real-world instances of an Argentine city demonstrate that the algorithm is competitive, obtaining, in the case of small instances, the same outcomes as those of an exact solver enhanced by valid inequalities, although requiring more computational time (as expected), and significantly improving the results of the latter for the case of larger instances, now requiring much less computational time. Thus, Scatter Search proves to be a competitive algorithm for addressing waste collection problems.

Keywords: municipal solid waste; waste collection; scatter search; mixed-integer programming; valid inequalities; case study

1. Introduction

In modern societies, municipal solid waste (MSW) management has emerged as a key priority for ensuring environmental sustainability, public health and the overall well-being of communities [1]. Among the various stages of the MSW system, the waste collection stage is one of the most tricky since it poses significant challenges due to its logistical complexity [2]. Implementing an efficient waste collection system is essential not only for maintaining the livability and quality of life for residents, but also for environmental protection and economic efficiency. Effective route management can significantly reduce operational costs and improve the overall efficiency of the waste management process [3] and reduce greenhouse emissions [4].

Moreover, route optimization and planning contributes to greater predictability in waste collection services. An efficient system allows residents to know when to expect collection, which helps to reduce waste accumulation in public areas and avoids the need for emergency services or additional collections outside scheduled times. This predictability minimizes the costs associated with emergency management and unforeseen interventions. Furthermore, improved planning can enhance the efficiency of scheduling other municipal services by coordinating routes to avoid conflicts with public transportation or emergency services, ultimately reducing delays and operational costs related to congestion and urban planning.

In Argentina, where the case study is focused, the design of waste collection routes is primarily based on empirical knowledge from decision-makers [5]. While this practical

experience is invaluable, the literature offers extensive insights into cost reduction and environmental impact through computational models that support the decision-making process [6–8]. By providing a solid foundation for decision-making, these computational support tools enable municipal managers to make adjustments and adapt the system to the changing needs of the city. However, waste collection problems are well-known NP-hard challenges, making them computationally difficult to solve [9]. Consequently, the design of efficient computational methods, based mainly on heuristic and metaheuristic approaches that can provide good solutions in reasonable computational times, are commonly employed in the literature.

This study proposes computational models to address the waste collection problem in the Argentine city of Bahía Blanca, focusing specifically on the downtown area. This sector was chosen because the collection points, where citizens will deposit their waste, have already been established. These points are strategically located to maximize waste management efficiency and facilitate access for both citizens and collection trucks [10]. The waste collection problem is mathematically modeled as a vehicle routing problem with time limits and capacity constraints. This problem is solved in two ways: exactly, using mixed-integer programming, and heuristically, using a Scatter Search (SS) algorithm. The exact resolution method is based on mixed-integer programming using different commercial solvers and valid inequalities to enhance the formulation. The metaheuristic approach employs an SS algorithm, which has proven efficient in solving various combinatorial optimization problems. As far as we know, this article is the first work to apply SS to the waste collection problem and to perform with this approach extensive computational experimentation using real-world instances. Considering different combination and improvement methods and different local search sizes, 36 SS configurations were implemented. Both exact and metaheuristic methods were compared in computational experiments based on real-world data from Bahía Blanca. The results show that SS is effective in tackling the problem, producing optimal solutions with small instances, although with longer computational times—as would be expected when comparing an exact method with a metaheuristic on problems easily tackled by the exact method. However, with larger instances, the SS algorithm clearly outperforms the exact method, which—even when enhanced by valid inequalities—struggled to handle the complexity of the instances. For larger instances, the SS algorithm obtains better results in smaller computing times.

In this context, this article contributes to the related literature with the following: (i) the first Scatter Search algorithm to address real-world instances of the waste collection problem in the literature; (ii) a study of the capacity to enhance the exact resolution of the problem using valid inequalities; and (iii) a computational experiment conducted on real-world instances from an Argentine city. This article is an invited extension of our conference work presented at the "VI Ibero-American Congress of Smart Cities" that was held from 13–17 November 2014 in Mexico and Cuernavaca City [11]. Compared to our conference work, the enhancements in this article include addressing a different optimization problem within waste management and introducing a new metaheuristic solution method. Additionally, we have expanded the literature review and computational experiments.

This article is structured as follows: Section 2 presents a review of the main related works. Section 3 develops the exact method for the waste collection problem, including the mathematical formulation and the used of valid inequalities to enhance the resolution process. Section 4 presents the Scatter Search algorithm and its main features. Section 5 describes the computational experimentation, including the description of the realistic instances, the tests performed over the valid inequalities for enhancing the exact resolution, the proposed implementations of the SS algorithm, and the comparison between the implementations of SS and the exact method. Finally, Section 6 outlines the main conclusion and future research lines.

2. Literature Review

The literature extensively addresses waste collection problems using various computational methods [12]. For a recent review on this topic, interested readers may refer to Hess et al. [2]. Waste collection problems are well-known variants of traditional vehicle routing problems, which are generally considered NP-hard [13]—that is, problems for which no efficient algorithms exist that can solve them in polynomial time with respect to the problem size. This class characterizes problems that are computationally difficult to solve [9]. As a result, heuristic and metaheuristic approaches have been widely used to tackle the computational complexity of waste collection problems [14]. While Scatter Search has been successfully employed to address several combinatorial optimization problems, including routing problems [15], it remains relatively underexplored in the context of waste management, in general, and waste collection problems, in particular [16].

According to a search performed in Scopus with the criteria schema TITLE-ABS-KEY (list of terms) using as keyword terms: ('waste' OR 'waste management') AND ('scatter search'), only five works related to the application of SS in waste management were found. Moreover, if the focus is put on waste collection, only two works are mentioned. One is the work of Zhang et al. [17]. Although the conceptual model in this work refers to the collection of recyclable waste, the SS algorithm was tested on synthetic instances built from the procedure developed in Dethloff [18]. The proposed SS outperformed a genetic algorithm in terms of computational efficiency, achieving similar results in significantly less time. Another related work is that of Chu et al. [19], who, while not applying SS directly to waste management, proposed waste collection as a potential application of their algorithm. Mainly they present an SS to address the periodic capacitated arc routing problem (PCARP). They compared SS with an insertion heuristic using PCARP instances derived from the benchmarks of Golden et al. [20] and Belenguer and Benavent [21], with SS outperforming the insertion heuristic in most cases.

Although no other applications of SS in waste collection routing problems were found, a few works have applied SS to other waste management-related problems. Yu et al. [22] used SS to optimize the transport of industrial waste from plants to intermediate recycling facilities and, eventually, to landfills using a transport-like mathematical model. They compared two SS versions: the typical SS and a modified SS where the update method is altered to maintain population diversity, rather than replacing the worst solution as in the traditional method. The modified version proved more efficient in tests conducted on synthetic, randomly generated instances.

Similarly, Ortega et al. [23] employed an SS hybridized with linear programming to design a multi-period reverse logistics network for MSW, which is mainly a location and transport optimization problem. In this network, the waste generated in towns has to be allocated to transfer centers and then to treatment plants. While the treatment plant location is fixed for the whole planning horizon, the location of the transfer centers may vary. The authors introduced randomness into the problem by varying waste generation based on empirical probabilistic distributions. They compared their algorithm with CPLEX. For small problems, SS reached the optimum in slightly more time than CPLEX, but for larger problems, SS was faster in reaching the optimum when CPLEX succeeded. The computational experiments were conducted on synthetic instances created by the authors, as well as on an instance partly based on the region of Gipuzkoa, Spain, and partly synthetic (mainly regarding waste generation).

Thus, the application of SS to real-world waste management problems is very limited. Most studies utilize synthetic instances that resemble waste management problems. Only the work of Ortega et al. [23] applied SS to a realistic instance for the network design of an actual region, although many key parameters were still synthetically generated. Concerning the specific problem of waste collection routing, the literature is even scarcer. Only two works address routing problems potentially representing waste collection, both of which rely on synthetic instances from benchmarks that are not related to waste management. Therefore, we consider that there is still a gap in the literature regarding the

implementation of Scatter Search—a metaheuristic that has proven efficient in solving other combinatorial optimization problems—specifically, for the waste collection problem, and its testing in realistic scenarios. As far as we know, this is the first work to propose an SS algorithm to the waste collection problem and perform an extensive computational experiment in real-world instances.

3. Exact Resolution

This Section presents the exact resolution approach used for the waste collection problem of Bahía Blanca. In particular, the basic mathematical formulation as a typical routing problem and a discussion about some valid cuts to improve the formulation are presented.

3.1. Mathematical Formulation

The typical waste collection problem consists in a set of collection points $I = \{1, 2, \ldots, n_I\}$ distributed in an area and a set of collection vehicles or trucks $V = \{1, 2, \ldots, n_V\}$. In addition, if 0 represents the depot where vehicles start and end their routes, and where collected waste is deposited, the set $I^0 = I \cup \{0\}$ is defined. The trucks have to depart from the depot, visit every collection point to obtain the waste and return to the depot to unload while respecting certain operative restrictions (truck capacity and travel time). The objective of the problem is to minimize the traveled distance, which is typically associated with reduced expenses, particularly in terms of fuel consumption and greenhouse gas emissions for diesel trucks like those used in the city of the case study. In order to present the model that is used in the exact solver, the following parameters are defined:

- Q: Truck capacity.
- c_{ig}: Travel time between points $i \in I^0$ and $g \in I^0$.
- se_i: Service time of collection point $i \in I^0$. At the depot, the service time corresponds to the time spent unloading the collected waste.
- b_i: Amount of waste generated daily at collection point $i \in I$.
- α: Cost per minute for collection trucks.
- T_L: Maximum time a vehicle can spend performing a route including unloading waste at the depot.

And the model uses the following variables:

- x_{igl}: (binary) 1 if truck $l \in L$ travels between collection points $i \in I^0$ and $g \in I^0$, and 0 otherwise.
- v_{igl}: (continuous non-negative) load of truck $l \in L$ when going from collection point $i \in I$ to collection point $g \in I$.

Taking all these elements into account, the following mathematical model based on Mixed-Integer Linear Programming (MILP) is proposed:

Minimize

$$f(x) = \alpha \sum_{l \in L} \sum_{i \in I^0} \sum_{g \in I^0} \left(x_{iglt} \left(c_{ig} + se_i \right) \right) \tag{1a}$$

subject to

$$\sum_{\substack{i \in I^0 \\ i \neq g}} \sum_{l \in L} x_{igl} = 1, \quad \forall\, g \in I, \tag{1b}$$

$$\sum_{\substack{i \in I^0 \\ i \neq g}} x_{igl} - \sum_{\substack{i \in I^0 \\ i \neq g}} x_{gil} = 0, \quad \forall\, g \in I^0, l \in L, \tag{1c}$$

$$\sum_{i \in I} x_{0il} \leq 1, \quad \forall\, l \in L, \tag{1d}$$

$$\sum_{\substack{i,g\in I^0\\ i\neq g}} x_{igl}\left(c_{ig}+se_i\right) \leq T_L, \quad \forall\, l \in L, \tag{1e}$$

$$v_{igl} \leq Q x_{igl}, \quad \forall\, i \in I^0, g \in I^0, l \in L, \tag{1f}$$

$$\sum_{\substack{i\in I^0\\ i\neq g}} v_{igl} + h_g \leq \sum_{\substack{i\in I^0\\ i\neq g}} v_{gil} + Q\left(1 - \sum_{\substack{i\in I^0\\ i\neq g}} x_{igl}\right), \quad \forall\, g \in I, l \in L, \tag{1g}$$

$$x_{igl} \in \{0,1\}, v_{igl} \geq 0, \quad \forall\, i \in I^0, g \in I^0, l \in L.$$

Equation (1a) represents the objective of minimizing the routing cost, which is based on the sum of the travel times and the service times (time to empty the collection points) for each route performed by the fleet of trucks. Equation (1b) presents the requirement that every collection point is visited by a collection truck. Equation (1c) requires that if a collection point is visited by a truck, this collection point is also left by the same truck. Equation (1d) requires that each collection truck only performs at most one route. Equation (1e) checks that the route does not exceed the maximum time that the vehicle can spend performing a route, which includes traveling between collection points, collecting waste of the collection pint (service time), and unloading waste at the depot in the end of the route. Equation (1f) ensures that the capacity of the trucks is not exceeded. Finally, Equation (1g) is the subtour elimination constraint, which also keeps track of the accumulated waste throughout the route.

3.2. Valid Cuts

The routing problems are computationally challenging problems [13], primarily due to the presence of many symmetric solutions. Two solutions are considered symmetric if they result in the same objective function value but differ in variable assignments [24]. For instance, one route can be performed by any truck of the homogeneous fleet and the solution incurs the same cost. Thus, the resolution solver has no mechanism to exclude one of them and this may have an impact in the efficiency of the resolution process.

To address this issue, Valid Inequalities (VIs) can be introduced into the model. A VI is a constraint that tightens the feasible region of the problem without eliminating any optimal solutions. For a thorough review of VIs in routing problems, the interested reader is referred to the work of Dror et al. [25]. In this paper, the performance for improving the mathematical formulation of Section 3.1 is tested with four typical VIs that were used in Mahéo et al. [24].

3.2.1. Furthest Visit

Without loss of generality, since each collection point can only be visited by one of the trucks in the fleet, the first truck is forced to be the one that visits the collection point furthest from the depot. These symmetric solutions are avoided using VI (2).

$$\sum_{i\in I} x_{igl} = 0, \quad \forall\, l \in L, l > 0, g \in I, g = \arg\max_{g\in I}\{c_{0g}\}. \tag{2}$$

3.2.2. Vehicles Are Ordered

It is mandated that a vehicle or truck with index l can only depart from the depot if the truck with index $l-1$ has already done so. In scenarios where not all available trucks are utilized, an unused truck could be swapped for one that is in use. These symmetric solutions are prevented by means of VI (3).

$$\sum_{\substack{g\in I\\ g\neq 0}} x_{0gl} \leq \sum_{\substack{g\in I\\ g\neq 0}} x_{0gl-1}, \quad \forall\, l \in L, l > 0. \tag{3}$$

3.2.3. Vehicle Has to Start Empty

A vehicle or truck must begin its route with an empty load. This avoids solutions with varying delivery plans. That is, if a truck completes its collection tour without reaching full capacity, an alternative solution could be considered where the truck starts with any load less than the remaining unused capacity. These symmetric solutions are avoided using VI (4).

$$\sum_{\substack{g \in I \\ g \neq 0}} v_{0gl} = 0, \quad \forall\, l \in L. \tag{4}$$

3.3. Solvers and Implementation Details

Exact solvers play a pivotal role in addressing complex mathematical combinatorial optimization problems, such as routing problems. In recent years, commercial solvers have significantly enhanced their capabilities to handle computationally intensive tasks, driven by a competitive landscape aimed at providing decision-makers with the most effective options.

There are several solvers available for mixed-integer programming problems. For a comprehensive list of solvers, refer to the NEOS Server website [26]. Among open-source solvers are CLP, CBC, LP_Solve, and GLPK [27]. In general, noncommercial MILP software tools may lack the speed and robustness of their commercial counterparts but provide a practical alternative for users who cannot afford commercial options [28]. For commercial solvers, a wide range is available, including CPLEX, Gurobi, XPRESS, and LINDO [29], with Gurobi and CPLEX among the most commonly used [30,31].

In line with this, the present study compares the performance of two renowned, state-of-the-art solvers: Gurobi version 11.0.3 [32] and CPLEX version 22.1.1 [33], both versions released in 2024. By evaluating these solvers, we aim to provide valuable insights for researchers and practitioners seeking to optimize complex combinatorial optimization problems.

The mathematical model and VIs presented in the previous section were implemented in Python using Pyomo [34] as a modeling environment. Pyomo provides a flexible platform for defining optimization problems and enables connections to various commercial solvers, including Gurobi and CPLEX, which were utilized in this study.

4. Scatter Search Algorithm

The Scatter Search algorithm—proposed by Glover in 1977 [35]—is a metaheuristic optimization technique aimed at solving complex combinatorial and continuous optimization problems. A core feature of Scatter Search is its focus on maintaining a reference set of diverse solutions, which are combined to generate new candidate solutions. This process enhances the likelihood of finding high-quality solutions [36]. In addition, the flexibility of SS allows adaptation to various problem domains, making it applicable in areas such as scheduling, routing, and resource allocation. By effectively balancing exploration and exploitation, SS aims to provide high-quality solutions efficiently. Its successful application across fields like logistics, telecommunications, finance, and engineering demonstrates its capability in addressing real-world optimization challenges [15]. As shown in Figure 1, it consists of five methods [37]:

- A diversification generation method used to generate a set of diverse solutions that are the basis for initializing the search. The outcome consists of a set of solutions that is referred to as the population.
- An improvement method for modifying solutions to improve their quality (in terms of the objective function) or feasibility.
- A reference set update method used in the main iterative loop of any scatter search implementation. The output is a set of solutions known as the reference set. Typically, the number of reference solutions is of the order of 10% of the population size.

- A subset generation method that produces subsets of reference solutions which become the input to the combination method.
- A combination method that uses the output from the subset generation method to create new solutions.

The main iterative loop is executed as long as at least one New Reference Solution (NRS) is obtained. The process continues until the Stopping Criterion is Satisfied (SCS).

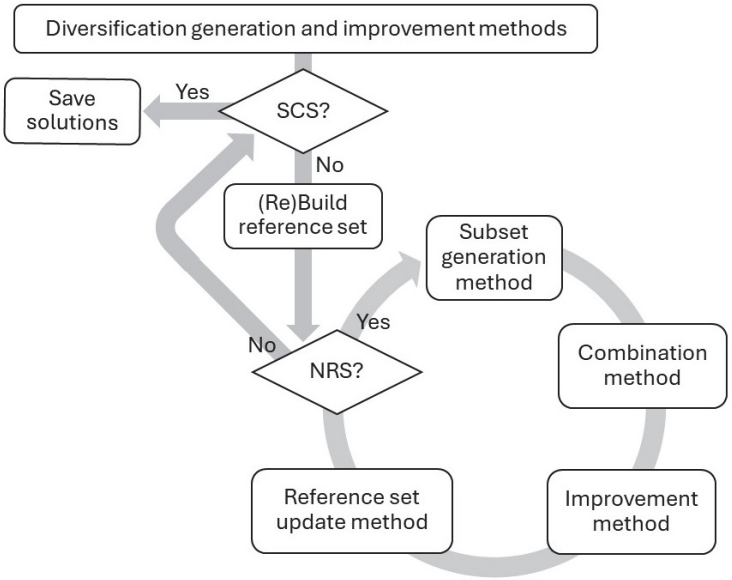


Figure 1. Scatter Search algorithm flowchart.

In this paper, the SS algorithm will be applied using various combinations of some key elements of the algorithm in order to evaluate the effectiveness of each configuration in addressing the waste collection problem. Specifically, different combination and improvement methods and different sizes of the local search, in the sense of the number of times the improvement method is applied to a solution, will be tested. This approach aims to identify the most effective configurations to optimize waste collection routes and improve overall system performance. The SS was also coded using Python as the programming language.

5. Computational Experimentation

This Section presents the description of instances considered, the tests performed on the VIs with the exact method, the proposed implementations of the SS algorithm and the comparison between the SS and the exact method.

5.1. Description of Instances and Resolution Platform

To carry out the computational experimentation in this work, 10 realistic instances were constructed with varying numbers of collection points: three instances with 15, three with 30, three with 50, and one with 100 collection points. All the instances are based on information gathered on field studies in Bahía Blanca [10]. These instances can be retrieved from Github (https://github.com/diegorossit/Urban_Science_waste_collection_BBCA.git, accessed on 27 October 2024). The naming convention follows the format: $n - id$, where n is the number of collection points, and id is a reference number to distinguish between instances with the same number of collection points.

Each instance consists of two files: "times.txt", which contains the matrix of travel times (minutes) from the depot to the collection points, between collection points, and from the collection points to the depot; and "waste.txt", which contains the geographic coordinates of each collection point, as well as the waste (m^3) generated daily by the urban area assigned to each collection point. Moreover, the service time of a collection

point was set at 0.78 min based on the field work of Carlos et al. [38] for a homogeneous collection vehicles fleet; the truck unloading time was set at 8 min [39], and the cost per minute of truck operation (α) was estimated at USD 0.5764 per minute [40]. Also, as in Mahéo et al. [24], given that the instances are smaller than the actual collection zones in the city, the capacity and size of the fleet were adjusted to ensure that the problem does not become trivial, where a single truck can collect all the waste in a single trip (see Table 1). In all cases, the maximum time to complete a trip or route was set at 6 h (the working hours of an employee in the sector in Bahía Blanca).

Table 1. Problem parameters.

Instance Size (Number of Collection Points)	Truck Fleet Size	Truck Capacity (m^3)
15	8	10
30	16	20
50	20	21
100	20	21

The resolution platform that was used for the SS and the exact method are a personal computer with a Processor Intel(R) Xeon(R) Silver 4216 CPU @ 2.10 GHz, 2095 Mhz (2 threads) and 64 GB of RAM within a Windows 11 operative system environment and Python compiler version 3.10.11.

5.2. Tests over VIs and Solver of Exact Method

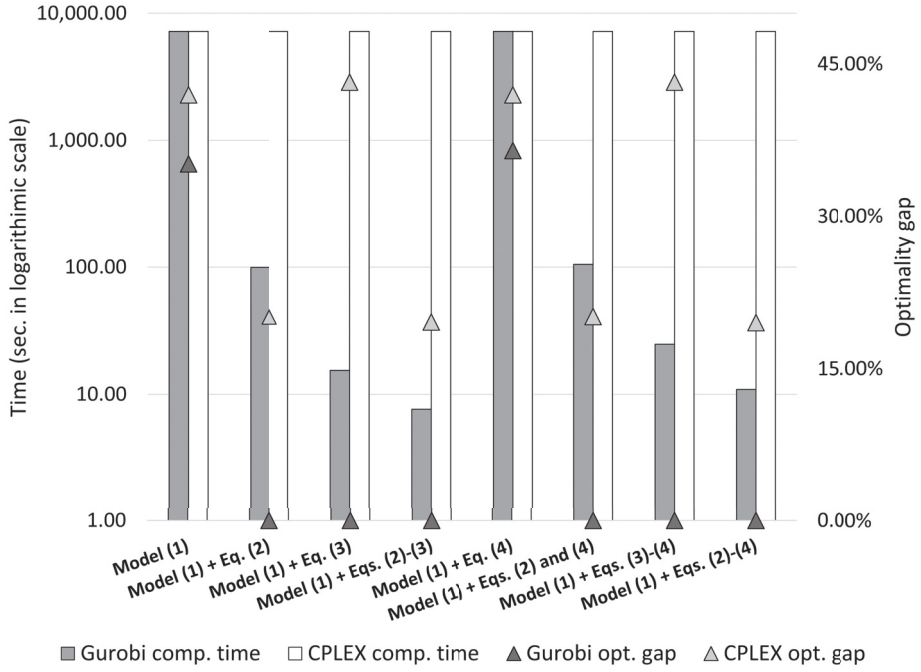
This section presents the results of tests analyzing the effect of VIs in enhancing the MILP model introduced in Section 3.1. As aforementioned, the tests were conducted using two state-of-the-art solvers, CPLEX and Gurobi, on the three smallest instances involving 15 collection points. Considering that there were three different VIs and all combinations between VIs and the MILP model were tested, eight runs were performed for each solver and each instance. A maximum allowable solving time of 7200 s was used for all cases.

Figure 2 illustrates the key results for the three instances. The bars show computing times (left x-axis), while triangle markers indicate optimality gaps (right x-axis), which are non-zero when the solver fails to converge within the allotted time (7200 s). For the sake of clarity, computing times are displayed on a logarithmic scale. The results of the detailed computing times and optimality gaps are presented in Appendix A. As aforementioned, each instance was tested with various mathematical formulations that resulted from the combination of Model (1) and the VIs from Equations (2)–(4). For example, the tag "Model (1) + Equations (2) and (4)" indicates that VIs (2) and (4) were added to Model (1). In particular, the tag "Model (1)" represents the base model presented in Section 3.1 without any VIs.

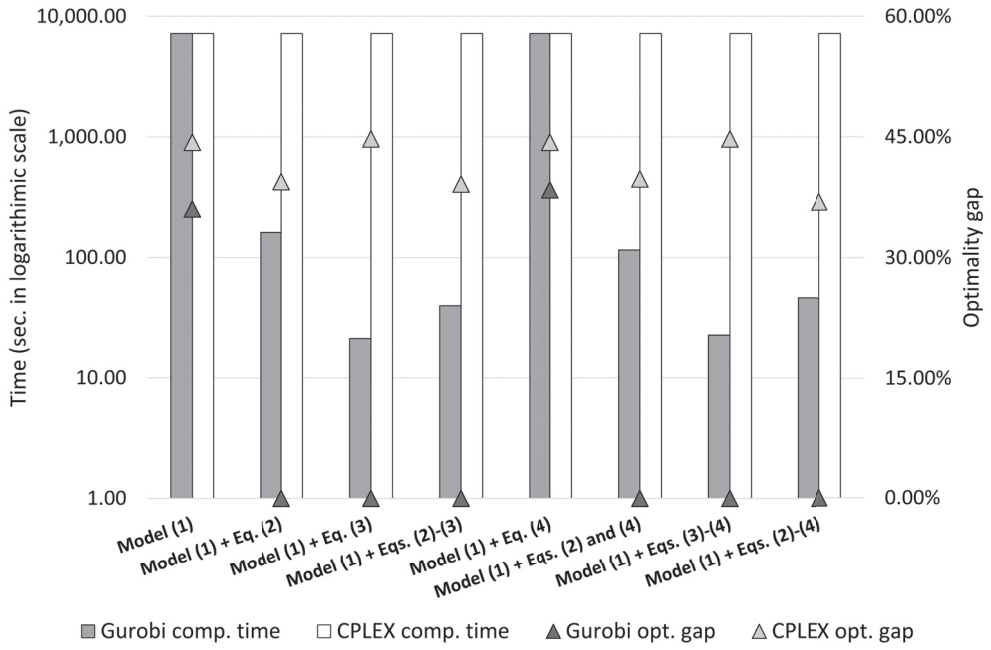
The comparison between the solvers reveals that Gurobi consistently outperforms CPLEX in these instances. In some cases, the difference is significant, such as in instance 15-1 with the formulation "Model (1) + Equations (3)–(4)", where Gurobi finds the optimal solution in under 8 s, while CPLEX only achieves a suboptimal solution with a 20% gap after reaching the time limit. Moreover, CPLEX fails to converge to the optimal solution for any instance or formulation. Additionally, in the formulations in which both solvers cannot obtain the optimal solution (e.g., "Model (1)" and "Model (1) + Equation (4)" across all three instances), Gurobi performs better, showing a smaller optimality gap.

Although the internal mechanisms of commercial solvers are not publicly accessible, these results align with recent findings where Gurobi outperformed CPLEX in complex integer problems. The detailed results are depicted in Appendix A, where it is shown that while CPLEX often finds the optimal solution, it typically fails to prove optimality within the time limit due to a weak lower bound. Nonetheless, as the literature emphasizes, Gurobi's performance advantage is problem-specific; different types of problems can lead to cases where either solver may outperform the other [30,41–43]. Thus, the recommendation for

Gurobi in this study is derived specifically from the computational experiments conducted here, without implying its superiority across all problem contexts.

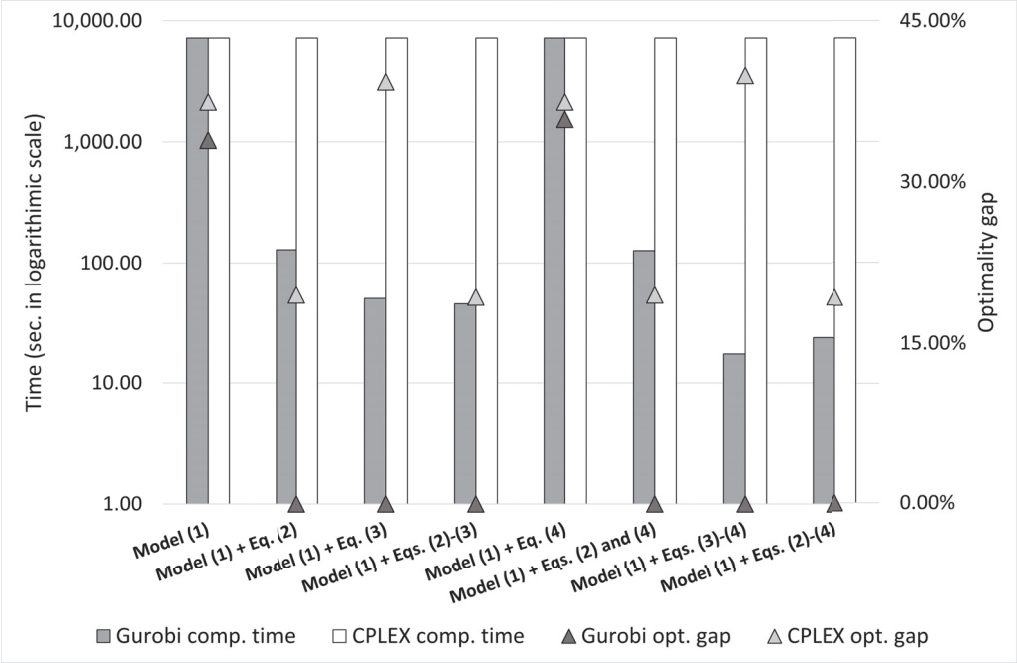


(**a**) Instance 15-1.



(**b**) Instance 15-2.

Figure 2. *Cont.*



(**c**) Instance 15-3.

Figure 2. Comparison of solvers and VIs for the exact method.

Regarding the effectiveness of VIs, in instance 15-1, the best result with Gurobi is achieved using formulation "Model (1) + Equation (2) and (3)" with a computing time of 7.63 s. For instances 15-2 and 15-3, the best results are obtained with the formulation "Model (1) + Equations (3) and (4)" yielding times of 22.63 s and 17.45 s, respectively. For the comparison with the Scatter Search algorithm, the Gurobi solver was used, applying the formulation "Model (1) + Equations (2) and (3)" as it achieved the smallest average computing time across the three tested instances.

5.3. Implementations of the SS Algorithm

This section presents the SS algorithms implemented to solve the problem described in Section 3.1. A permutation coding was considered for the representation of the solutions. Thus, the variable x is replaced by the chromosome $y = [y_i]$, with a permutation of the order of collection of waste from the n_I collection points and the variable v is determined when the fitness function is calculated. Thus, by handling permutations, constraint (1c) is always satisfied. The remaining constraints are imposed when evaluating the fitness function. On the one hand, restrictions (1b), (1d), (1f), and (1g) are met by adding trucks to the overall route (defined by the chromosome) once the previous truck has filled up or does not have the capacity to load all the waste deposited at the next collection point. On the other hand, solutions are penalized in the event they do not meet restriction (1e) or the size of the fleet (n_L) is exceeded, as shown in Equation (5).

$$f_{fitness}(x) = f(x) - \lambda\, min\{0, n_L - n(x)\} - \gamma \sum_{l \in L} min\{0, T_L - TT_l\} \quad (5)$$

where $f(x)$ is the objective function (1a), λ and γ are positive real numbers, $n(x)$ is the number of trucks required to collect all the accumulated waste, one for each route of the solution x, and TT_l is the summation of the left-hand side of the constraint (1e).

The following are the different methods that were considered:

- Diversification generation method: more_intertools.random_permutation() (https://pypi.org/project/more-itertools/, accessed on 27 October 2024).
- Improvement methods:

1 Exchange (EXC). Two alleles are chosen randomly and exchanged with each other.
2 Insertion (INS). Two alleles are chosen randomly and the second allele is placed just after the first one.
3 Inversion (INV). Two alleles are chosen at random and the order of the alleles between them is reversed.

- Reference set update method: the first half of the solutions in the reference set are chosen to be the best solutions (in terms of the fitness function value) in the population. The other half are chosen to be the most diverse (in terms of the Hamming distance) with respect to the solutions already incorporated in the reference set.
- Subset generation method: all possible pairs of solutions of the reference set are generated.
- Combination method:

1 Partially Mapped Crossover (PMX) [44].
2 Order Crossover (OX) [45]).
3 Circle Crossover (CX) [46].
4 Modified Circle Crossover (CX2) [47].

In addition, three sizes for local search were considered for the improvement methods: (1) 10, (2) 20, and (3) 30. Thus, in total, 36 configurations of the SS algorithm were made. Since SS is based on stochastic procedures, for each instance and SS configuration, 31 runs were made to obtain a proper sample of the distribution of the results for statistical analysis. A maximum number of evaluations of the fitness function was set as a stopping criterion for the algorithms. Table 2 shows the size of the reference set, the size of the population and the maximum number of fitness function evaluations considered according to the instance size.

Table 2. Algorithm parameters.

Instance Size (Number of Collection Points)	Reference Set Size	Population Size	Maximum Number of Fitness Function Evaluations
15	10	90	100,000
30	10	90	250,000
50	12	132	500,000
100	14	182	1,000,000

For instances of 15 and 30 collection points, a Kruskal–Wallis Rank Sum Test [48] was performed to determine whether there are significant differences between the four combination methods considered, the three improvement methods considered, and the three sizes for local search considered. When the value of a Kruskal–Wallis test is significant (p-value < 0.05), a multiple comparison test after Kruskal–Wallis [49] between treatments was performed to determine which levels are different, with pairwise comparisons adjusted appropriately for multiple comparisons. The tests were performed using the available libraries in the programming language R [50].

Figures 3 and 4 show the box-and-whisker plots associated with the results of the instances of 15 and 30 collection points, respectively, grouped by combination methods, improvement methods, and local search sizes. In both cases, it can be seen that the PMX and OX combination methods show the best performance, as well as the EXC improvement method. The size of the local search does not seem to be relevant. These conclusions are supported by the hypothesis tests performed (see Tables 3 and 4). A detailed statistical summary of the computational executions can be found in Appendix B.

Table 3. Kruskal–Wallis Rank Sum Test: *p*-values.

Instance	Combination Method	Improvement Method	Local Search Sizes
15-1	2.2×10^{16}	1.281×10^{08}	0.9543
15-2	2.2×10^{16}	6.430×10^{-11}	0.0799
15-3	2.2×10^{16}	5.504×10^{11}	0.8066
30-1	2.2×10^{16}	2.2×10^{16}	0.7939
30-2	2.2×10^{16}	2.2×10^{16}	0.2394
30-3	2.2×10^{16}	2.2×10^{16}	0.3931

Table 4. Outputs of the multiple comparison test after Kruskal–Wallis: if TRUE, then statistically significant differences are found between the compared levels.

Combination Method		Improvement Method
1–2 FALSE	2–3 TRUE	1–2 TRUE
1–3 TRUE	2–4 TRUE	1–3 TRUE
1–4 TRUE	3–4 TRUE	2–3 TRUE

Note: The outputs are the same for all instances of 15 and 30 collection points.

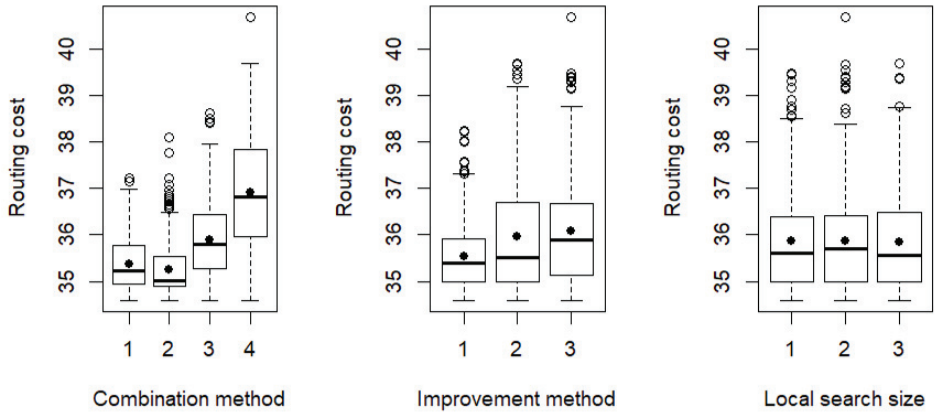
Taking into account the conclusions obtained with the instances of 15 and 30 collection points, the larger instances with 50 and 100 collection points were run with only the configurations that were more efficient, i.e., the configurations that use PMX or OX as the combination method and EXC as the improvement method, with a local search size of 20. Table 5 shows the statistical summary of the configurations executed on instances 50-1, 50-2, 50-3, and 100-1. From left to right, the table reports: the minimum (the best value of the objective function found), the first quartile, the median, the second quartile, the (pseudo) median, the lower bound and the upper bound of a 95% confidence interval for the (pseudo) median, the total runtime (in seconds), of the 31 runs performed in each case, and the mean runtime of each run (in seconds).

Table 5. Statistical summary (runtimes in seconds) of the configurations executed with the 50-1, 50-2, 50-3, and 100-1 instances.

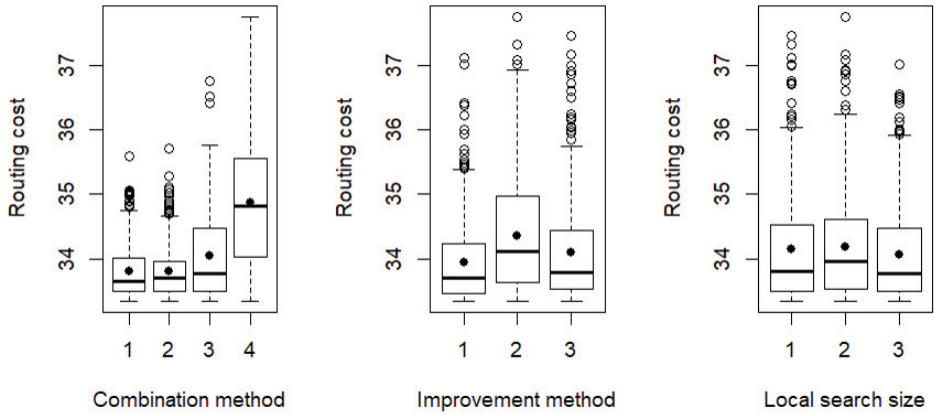
Instance	Comb. Method	min	Q1	Median	Q2	(Pseudo) Median	lb	ub	Total Runtimes	Mean Runtimes
50-1	PMX	79.36150	83.38203	84.75679	86.69644	84.96863	84.06797	85.97016	2222.0133	71.6778
	OX	82.68168	83.97286	85.37932	86.59845	85.30007	84.68186	85.97304	2285.4383	73.7238
50-2	PMX	83.75382	86.82326	87.46020	88.33636	87.44579	86.90684	88.11156	2217.8015	71.5420
	OX	83.06212	86.43129	87.73112	88.63898	87.62016	86.86361	88.21819	2296.5859	74.0834
50-3	PMX	82.61251	84.89225	85.82317	87.49190	86.16902	85.42832	86.90107	2219.7335	71.6043
	OX	81.53460	83.37627	85.11994	87.16335	85.19631	84.36194	86.14020	2290.5435	73.8885
100-1	PMX	151.6964	158.0572	159.8413	162.7925	160.0127	158.6192	161.2967	4863.0509	156.8726
	OX	156.4116	158.7144	159.7951	161.5936	159.9803	159.1870	160.8702	5184.1772	167.2315

Note: The 100-1 instance was run in a Processor 13th Gen Intel(R) Core(TM) i9-13900K with 128 GB of RAM.

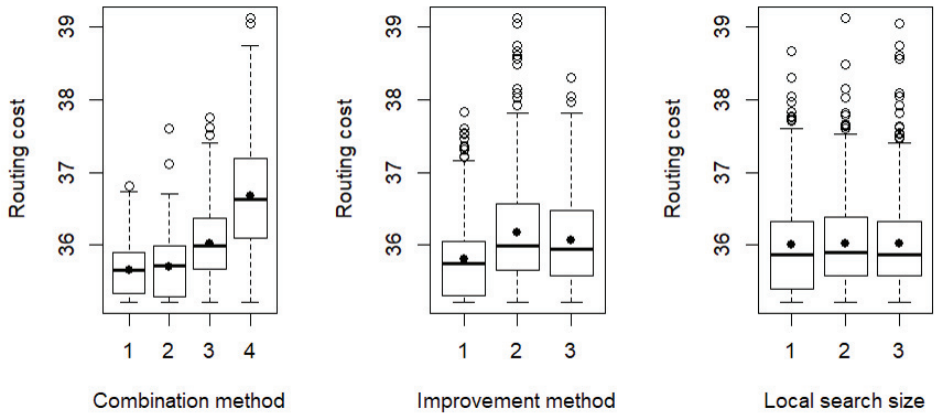
The results of Table 5 show that the mean and total runtime increase approximately linearly with the size of the problem, which is a good result of an NP-hard problem.



(**a**) Instance 15-1.

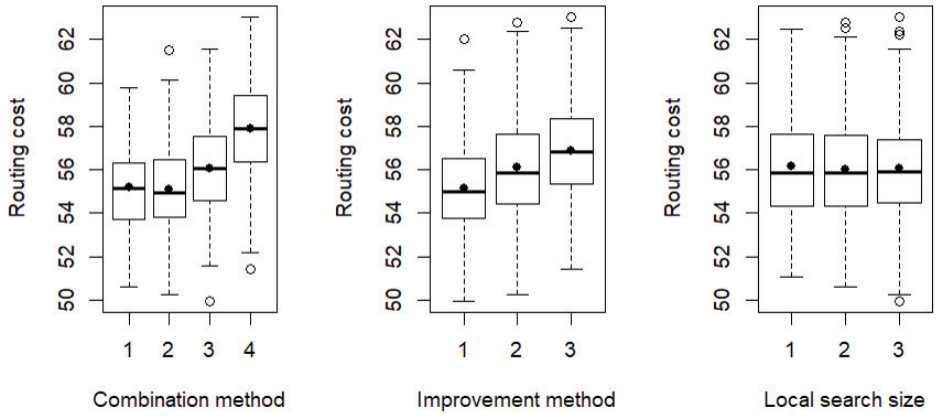


(**b**) Instance 15-2.

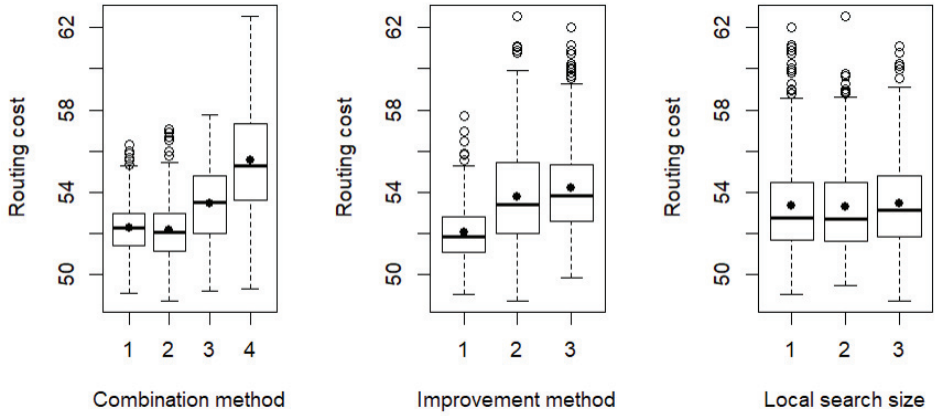


(**c**) Instance 15-3.

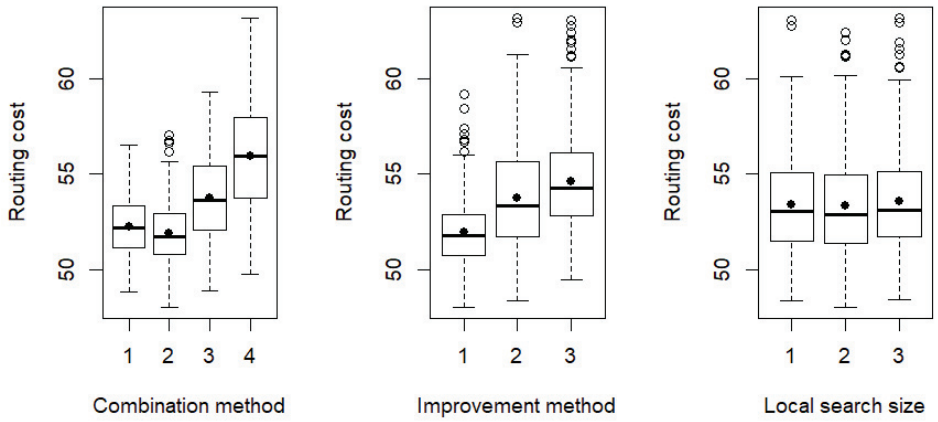
Figure 3. Box-and-whisker plots associated with the results of the three instances of 15 collection points grouped by combination methods, improvement methods, and local search sizes.



(**a**) Instance 30-1.



(**b**) Instance 30-2.



(**c**) Instance 30-3.

Figure 4. Box-and-whisker plots associated with the results of the three instances of 30 collection points grouped by combination methods, improvement methods, and local search sizes.

5.4. Comparison Between the Scatter Search and the Exact Method

In this section, we compare the results obtained with the Gurobi solver, applying the formulation "Model (1) + Equations (3) and (4)", with the SS algorithm configuration that uses OX as the combination method, EXC as the execution method, and 20 as the

size of the local search, since it is statistically one of the most efficient among the tested SS configurations. Table 6 shows the routing cost, lower bound, computing time in seconds, and optimality gap of the exact method; and the percentage difference, calculated with Equation (6), between the routing cost obtained with the exact method and the minimum and (pseudo) median values obtained with the (OX, EXC, 20) configuration of the SS algorithm.

$$\%\ \text{dif.} = \left(\frac{C_*^{SS} - C_*^{exact}}{C_*^{exact}} \right) \% \tag{6}$$

In Equation (6), C_*^{SS} is the routing cost obtained with the SS and C_*^{exact} is the routing cost obtained with the exact method. Thus, a negative value of % dif. means that the SS algorithm obtained a smaller (better) solution. In Table 6, the Equation (6) was calculated using two different values for C_*^{SS}: the minimal (best) value and the median value obtained for the 31 runs for each instance.

From the results presented in Table 6, it can be concluded that for the instances with 15 collection points, the SS algorithm was able to consistently obtain optimal solutions. Moreover, the median result across the 31 runs of the SS algorithm showed less than a 2% deviation from the optimal value for all three instances with 15 collection points. This validates the SS algorithm for small instances, as it is capable of producing near-optimal solutions in most runs. The total runtimes of the SS are much larger (around 250 s; see Tables A2–A4) than the exact method (around 20 s) as is expected for small instances in which the stochastic search of metaheuristics is overrun by the systematic search of the exact method. However, the mean runtime of the SS is competitive against the exact method (around 7 s, see Tables A2–A4).

Table 6. Comparison of the results obtained with the exact method and the SS algorithm.

Instance	Exact Method				Exact Method vs. SS	
	Routing Cost (USD)	Comp. Time (s)	Lower Bound	Opt. Gap (%)	min (%dif)	(Pseudo) Median (%dif)
15-1	34.59	24.80	34.59	0.00%	0.00%	1.99%
15-2	33.35	22.63	33.35	0.00%	0.00%	1.80%
15-3	35.22	17.45	35.22	0.00%	0.00%	1.00%
30-1	67.57	7201.82	34.78	94.27%	−23.41%	−19.37%
30-2	80.61	7200.75	35.16	129.27%	−38.49%	−36.05%
30-3	48.16	7201.73	35.11	37.17%	−00.25%	5.56%
50-1	123.57	7201.58	49.87	147.78%	−33.09%	−30.97%
50-2	-	7200.03	51.81	-	-	-
50-3	-	7200.02	49.96	-	-	-

Note: The minimum and (pseudo) median values are taken from the configuration (OX, EXC, 20).

For instances with 30 collection points, the SS algorithm clearly outperforms the exact method in terms of finding the best minimal solution. The routing plans generated by the SS algorithm were more than 20% better than those produced by the exact method for instances 30-1 and 30-2. However, in instance 30-3, the percentage difference between the SS and the exact method was much smaller, which aligns with the exact method's ability to achieve a relatively small optimality gap for this particular instance size. The median results follow a similar pattern, with the SS algorithm significantly outperforming the exact method in instances 30-1 and 30-2, while in instance 30-3, the SS median result was about 5% worse than the exact method's solution. In these instances, the total computing times of the SS are much smaller (around 700 s, see Tables A5–A7) than for the exact method, which used the maximum allowable time of 7200 s for all three instances. The exponential increase in computational time for the exact method as the instance size doubled from 15 to 30 collection points is typical of NP-hard problems.

For the larger instances with 50 collection points, only the instance 50-1 could be compared, as the exact method failed to find a solution within the allowable time (7200 s)

for the other two instances. In instance 50-1, the SS algorithm improved the exact solution by 33.09% when considering the best solution and by 30.97% when considering the median solution.

These results highlight the robustness of the SS algorithm in larger instances, where it clearly outperforms the exact method in both solution quality and computational efficiency.

6. Conclusions

Municipal Solid Waste (MSW) systems are critical components of modern societies, playing a crucial role in enhancing the sustainability and livability of cities. To achieve these objectives, various stages of the system must be carefully planned. This article addresses the waste collection problem, a complex logistics challenge that is often computationally expensive to solve to optimality. Waste collection is one of the most costly stages of the MSW system. Therefore, implementing computational intelligence methods to optimize this stage, which is typically difficult to solve manually, can reduce costs and significantly impact the system's overall expenses. In this line, this work proposes a Scatter Search algorithm to tackle the problem efficiently, being the first work in the literature to apply this approach to a real-world case study of the waste collection problem. The case study focuses on the medium-sized Argentine city of Bahía Blanca. Additionally, an exact resolution of the problem by means of mathematical programming is proposed.

Computational experimentation was performed over scenarios of different sizes to study the impact of the computational complexity. The exact method was tested using different valid inequalities to enhance the formulation and two different state-of-the-art commercial solvers. The results showed that Gurobi was significantly more efficient than CPLEX for this problem, and that the valid inequalities greatly strengthened the formulation, allowing the model to obtain the optimal solution in small instances, something that was not possible for the plain mathematical model without the valid inequalities. The Scatter Search algorithm was tested using different configurations for the improvement method, the combination method, and the size in the local search operator, giving a total of 36 different configurations of the Scatter Search algorithm. The results showed that the configurations using partially mapped crossover or order crossover as the combination method and exchange as the improvement method are the most efficient to address this problem. Moreover, the results showed that the size of the local search is not so relevant for the target problem. Finally, the comparison between the exact method with the best formulation and the Scatter Search revealed that the heuristic approach consistently found near-optimal solutions in smaller instances. Additionally, in larger instances, the SS was able to find solutions that outperformed the exact method with much smaller computing times. Overall, the experimentation demonstrated the competitiveness of the Scatter Search algorithm in addressing large real-world scenarios in the waste collection problem. Additionally, it confirmed the ability of valid inequalities to enhance the formulation model, facilitating its application to more complex and realistic instances of waste collection challenges.

Future research lines could explore hybrid approaches that combine Scatter Search with other metaheuristics and perform a comparative analysis between Scatter Search and other computational intelligent mataheuristics. Additionally, dynamic models can be applied for real-time adjustments to planned routes based on smart bins information, integrating smart city technologies like IoT to enhance decision-making. Another critical research line is to incorporate sustainability assessments to evaluate the environmental impacts of collection routes alongside travel costs, thereby addressing the waste collection problem in a multi-objective manner. Finally, engaging with practitioners to discuss these solutions and gather feedback will be essential to ensure the practicality and effectiveness of the proposed strategies.

Author Contributions: Conceptualization, D.R., B.G.L., M.F. and M.M.B.; methodology, D.R. and B.G.L.; software, D.R. and B.G.L.; validation, D.R. and B.G.L.; formal analysis, D.R. and B.G.L.; investigation, D.R. and B.G.L.; resources, D.R. and B.G.L.; data curation, D.R. and B.G.L.; writing—original draft preparation, D.R., B.G.L. and M.F.; writing—review and editing, D.R., B.G.L., M.F. and M.M.B.; visualization, D.R. and B.G.L.; supervision, D.R., B.G.L., M.F. and M.M.B.; project administration, D.R., B.G.L., M.F. and M.M.B.; funding acquisition, D.R., B.G.L. and M.M.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Consejería de Economía, Industria, Comercio y Conocimiento of the Government of the Canary Islands through the direct grant awarded to the ULPGC called "Support for R+D+i activity. Campus of International Excellence CEI CANARIAS-ULPGC", by CONICET under research project PIBAA 0466CO, by the Agencia I+D+i of Argentina under research project PICT-2021-I-INVI00217, and by 319RT0574—Red iberoamericana Industria 4.0 of CYTED.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Instances can be retrieved from Github (https://github.com/diegorossit/Urban_Science_waste_collection_BBCA.git, accessed on 27 October 2024). Other datasets are available on request from the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Comparison Between Solvers and Valid Inequalities for Exact Method

This Appendix presents the detailed results of the comparison between the commercial solvers Gurobi and CPLEX and valid inequalities for the exact method. In Table A1, the results for each instance and solver are presented. For each instance, the following data are outlined: Upper Bound (UB) (i.e., the incumbent integer feasible solution obtained by the solver), computing time in seconds, optimality gap, and the lower bound (LB) (i.e., solution of the relaxed problem). The optimality gap is estimated as in Gurobi Optimization, LLC [32] with the information of the UB and LB with Equation (A1).

$$opt.\ gap = \left(\frac{UB - LB}{LB}\right)\% \tag{A1}$$

From the detailed results, it can be observed that CPLEX occasionally reaches the optimal solution but lacks sufficient information to confirm this due to a weak lower bound. In contrast, a key strength of Gurobi lies in its ability to generate competitive lower bounds, which significantly reduces computing times. This capability enables Gurobi to converge more effectively on optimal solutions, enhancing its performance relative to CPLEX.

Table A1. Detailed results for the comparison between solvers and VIs for the exact method.

Gurobi												
	15_1				15_2				15_3			
Formulations	**UB**	**Comp. Time (sec)**	**opt. gap**	**LB**	**UB**	**Comp. Time (sec)**	**opt. gap**	**LB**	**UB**	**Comp. Time (sec)**	**opt. gap**	**LB**
Model (1)	34.93	7203.08	35%	22.64	33.37	7204.15	36%	21.35	35.31	7203.75	34%	23.33
Model (1) + Equation (2)	34.59	99.20	0%	34.59	33.35	161.67	0%	33.35	35.22	129.31	0%	35.22
Model (1) + Equation (3)	34.59	15.40	0%	34.59	33.35	21.23	0%	33.35	35.22	50.60	0%	35.22
Model (1) + Equations (2) and (3)	34.59	7.63	0%	34.59	33.35	39.64	0%	33.35	35.22	45.59	0%	35.22
Model (1) + Equation (4)	34.93	7203.70	36%	22.19	33.52	7204.27	38%	20.65	35.71	7203.84	36%	22.91
Model (1) + Equations (2) and (4)	34.59	104.81	0%	34.59	33.35	115.65	0%	33.35	35.22	126.38	0%	35.22
Model (1) + Equations (3) and (4)	34.59	24.80	0%	34.59	33.35	22.63	0%	33.35	35.22	17.45	0%	35.22
Model (1) + Equations (2) and (4)	34.59	10.91	0%	34.59	33.35	46.10	0%	33.35	35.22	23.84	0%	35.22
CPLEX												
Formulations	**UB**	**Comp. Time (sec)**	**opt. gap**	**LB**	**UB**	**Comp. Time (sec)**	**opt. gap**	**LB**	**UB**	**Comp. Time (sec)**	**opt. gap**	**LB**
Model (1)	34.59	7203.66	42%	20.08	33.35	7203.32	44%	18.57	35.22	7209.65	37%	22.03
Model (1) + Equation (2)	34.59	7208.26	20%	27.60	33.35	7203.53	39%	20.21	35.22	7205.12	19%	28.36
Model (1) + Equation (3)	35.10	7203.38	43%	19.87	33.49	7206.26	45%	18.44	35.22	7206.83	39%	21.37
Model (1) + Equations (2) and (3)	35.10	7210.45	20%	28.12	35.16	7205.62	39%	21.34	35.22	7213.45	19%	28.43
Model (1) + Equation (4)	34.59	7203.27	42%	20.08	33.35	7204.08	44%	18.57	35.22	7203.04	37%	22.03
Model (1) + Equations (2) and (4)	34.59	7208.77	20%	27.62	33.64	7203.77	40%	20.20	35.22	7210.53	19%	28.36
Model (1) + Equations (3) and (4)	35.10	7202.95	43%	19.87	33.49	7205.99	45%	18.45	35.69	7209.10	40%	21.38
Model (1) + Equations (2) and (4)	35.10	7210.04	20%	28.14	33.93	7205.62	37%	21.33	35.22	7205.89	19%	28.42

Appendix B. Scatter Search. Statistical Summary of the Computational Experiments

This appendix contains the box-and-whisker plots of the 36 SS algorithm configurations considered with the instances of 15 (see Figures A1–A3) and 30 (see Figures A4–A6) collection points, and the corresponding statistical summary tables. Thus, Tables A2–A4 show the statistical summary of the experiments performed with the 15-1, 15-2, and 15-3 instances, respectively, and Tables A5–A7 show the statistical summary of the experiments performed with the 30-1, 30-2, and 30-3 instances, respectively. In each table, from left to right is reported, for each configuration of the SS algorithm: the minimum (the best value of the objective function found), the first quartile, the median, the second quartile, the (pseudo) median, the lower bound and the upper bound of a 95% confidence interval for the (pseudo) median, the total runtime (in seconds), of the 31 runs performed in each case, and the mean runtime of each run (in seconds). The pseudomedian of a distribution F is the median of the distribution of $(u + v)/2$, where u and v are independent, each with [48].

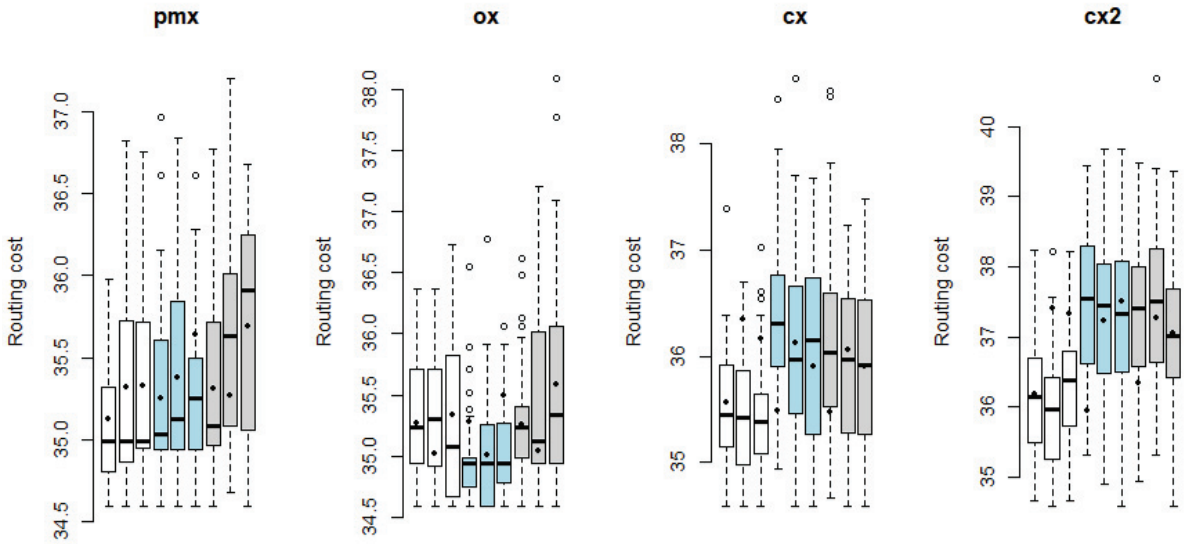


Figure A1. Instance 15-1. Box-and-whisker plots of the results grouped by combination methods, improvement method: EXC (white), INS (blue), and INV (gray), and local search sizes: 10, 20, and 30, from left to right.

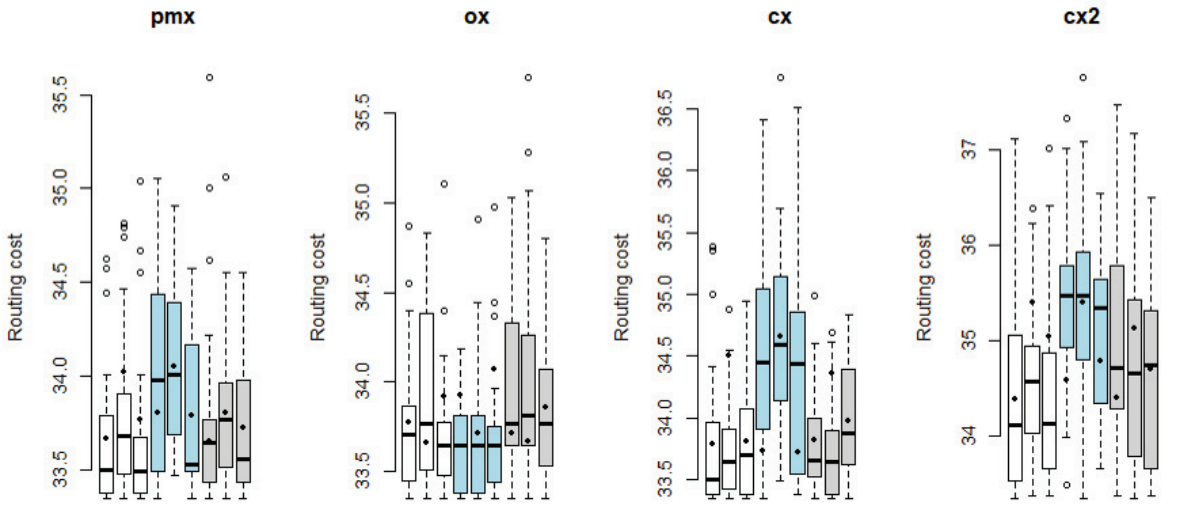


Figure A2. Instance 15-2. Box-and-whisker plots of the results grouped by combination methods, improvement method: EXC (white), INS (blue), and INV (gray), and local search sizes: 10, 20, and 30, from left to right.

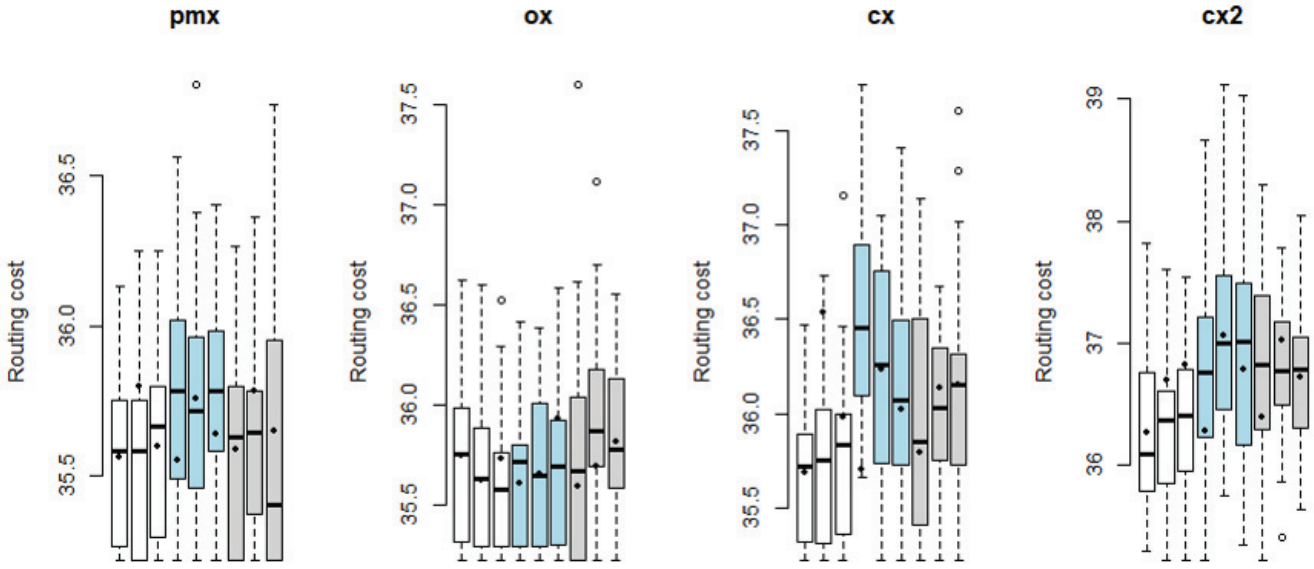


Figure A3. Instance 15-3. Box-and-whisker plots of the results grouped by combination methods, improvement method: EXC (white), INS (blue), and INV (gray), and local search sizes: 10, 20, and 30, from left to right.

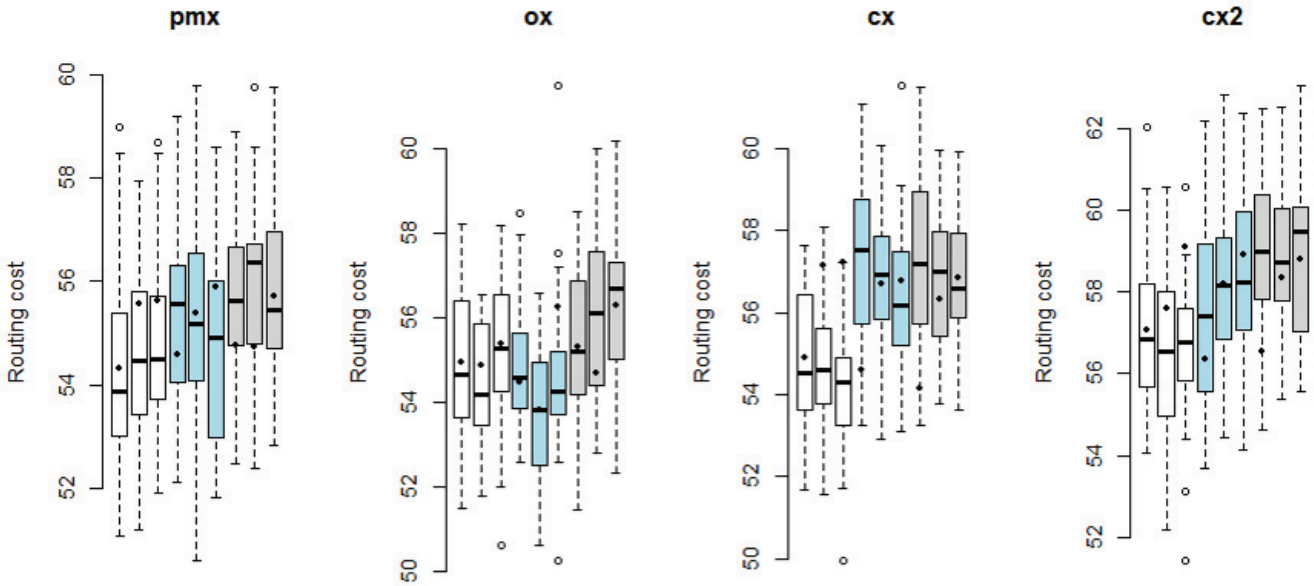


Figure A4. Instance 30-1. Box-and-whisker plots of the results grouped by combination methods, improvement method: EXC (white), INS (blue), and INV (gray), and local search sizes: 10, 20, and 30, from left to right.

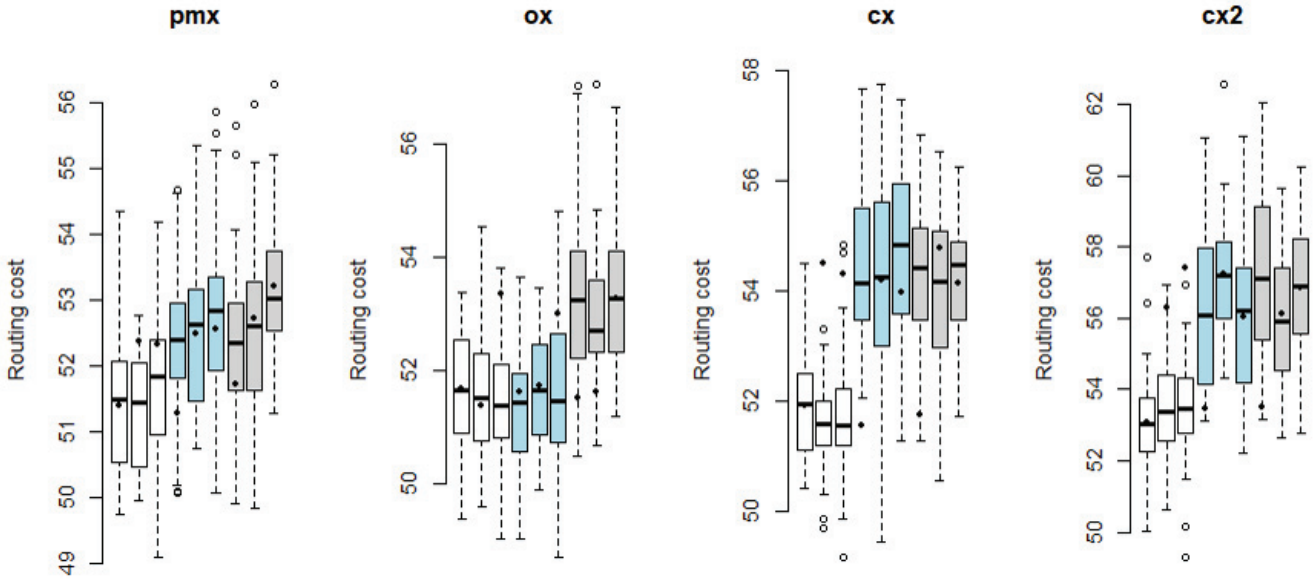


Figure A5. Instance 30-2. Box-and-whisker plots of the results grouped by combination methods, improvement method: EXC (white), INS (blue), and INV (gray), and local search sizes: 10, 20, and 30, from left to right.

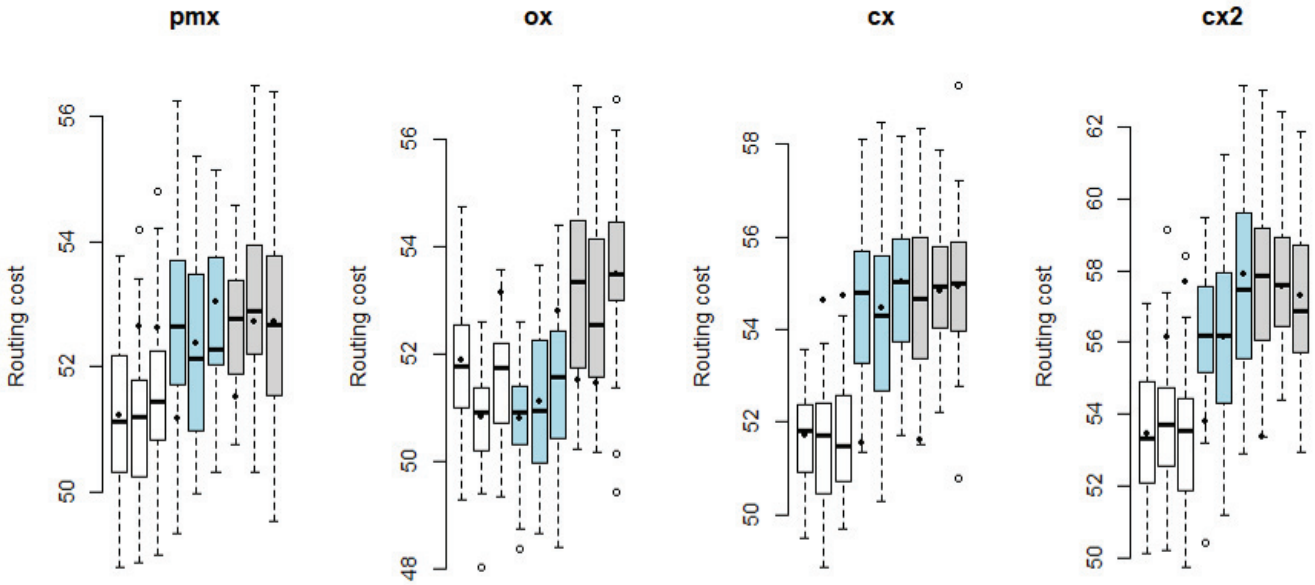


Figure A6. Instance 30-3. Box-and-whisker plots of the results grouped by combination methods, improvement method: EXC (white), INS (blue), and INV (gray), and local search sizes: 10, 20, and 30, from left to right.

Table A2. Instance 15-1. Statistical summary (runtimes in seconds).

Combin. Method	Improv. Method	Local Search Size	min	Q1	Median	Q2	(Pseudo) Median	lb	ub	Total Runtimes	Mean Runtimes
PMX	EXC	10	**34.59096**	34.80424	34.98869	35.31725	35.10107	34.95119	35.31720	243.68339	7.86075
PMX	EXC	20	**34.59096**	34.86188	34.98869	35.72651	35.21642	34.95989	35.40944	207.04524	6.67888
PMX	EXC	30	**34.59096**	34.94258	34.98869	35.71786	35.24804	34.98868	35.49298	191.32109	6.17165
PMX	INS	10	**34.59096**	34.93681	35.02904	35.60258	35.23934	35.00881	35.49601	263.45977	8.49870
PMX	INS	20	**34.59096**	34.93681	35.12703	35.83891	35.35184	35.08669	35.59676	229.96111	7.41810
PMX	INS	30	**34.59096**	34.93681	35.24808	35.49594	35.23083	35.04350	35.43840	217.88484	7.02854
PMX	INV	10	**34.59096**	34.95987	35.08092	35.71498	35.30537	35.03196	35.52474	267.14046	8.61743
PMX	INV	20	34.67166	35.08092	35.62852	36.01472	35.59175	35.37194	35.85050	231.33146	7.46231
PMX	INV	30	**34.59096**	35.05786	35.91096	36.25105	35.67753	35.44991	35.97443	219.01970	7.06515
OX	EXC	10	**34.59096**	34.93681	35.23079	35.70922	35.25384	35.07809	35.44986	231.14924	7.45643
OX	EXC	20	**34.59096**	34.91952	35.30572	35.71498	35.27983	35.08385	35.48148	194.77790	6.28316
OX	EXC	30	**34.59096**	34.67166	35.08092	35.82450	35.23946	34.95993	35.48443	182.80096	5.89681
OX	INS	10	**34.59096**	34.74660	34.93681	34.98581	34.95990	34.80425	35.13281	264.53177	8.53328
OX	INS	20	**34.59096**	**34.59096**	34.93681	35.25673	**34.93688**	**34.78975**	**35.10975**	229.01960	7.38773
OX	INS	30	**34.59096**	34.78694	34.93681	35.26826	34.96274	34.90220	35.15868	216.53589	6.98503
OX	INV	10	**34.59096**	34.98581	35.23079	35.40371	35.24462	35.10683	35.52475	267.98669	8.64473
OX	INV	20	**34.59096**	34.93681	35.12703	36.02048	35.45007	35.04638	35.79576	232.05914	7.48578
OX	INV	30	**34.59096**	34.93681	35.33454	36.06372	35.46722	35.15582	35.83898	219.81882	7.09093
CX	EXC	10	**34.59096**	35.15585	35.44983	35.91961	35.52714	35.32585	35.74099	222.05486	7.16306
CX	EXC	20	**34.59096**	34.98293	35.42101	35.86773	35.45504	35.22783	35.72363	189.66683	6.11828
CX	EXC	30	**34.59096**	35.08092	35.38642	35.64005	35.41524	35.22788	35.62569	178.91450	5.77144
CX	INS	10	34.93681	35.90520	36.30869	36.76118	36.33454	36.02624	36.65452	258.99309	8.35462
CX	INS	20	**34.59096**	35.45847	35.97437	36.66031	36.05195	35.73220	36.42977	224.35371	7.23722
CX	INS	30	**34.59096**	35.26249	36.15306	36.73236	36.07049	35.68336	36.42406	212.69208	6.86103
CX	INV	10	34.67166	35.52476	36.03201	36.59114	36.08384	35.78712	36.43552	259.40189	8.36780
CX	INV	20	**34.59096**	35.27402	35.96861	36.54503	35.91387	35.59974	36.23950	226.78883	7.31577
CX	INV	30	**34.59096**	35.26826	35.91673	36.52773	35.90070	35.57952	36.19918	216.42255	6.98137
CX2	EXC	10	34.67166	35.49594	36.15306	36.69489	36.18476	35.85039	36.49609	232.64759	7.50476
CX2	EXC	20	**34.59096**	35.26537	35.96861	36.42974	35.89079	35.65446	36.24241	196.65155	6.34360
CX2	EXC	30	34.67166	35.72939	36.38939	36.79577	36.28275	36.00893	36.66612	185.13147	5.97198
CX2	INS	10	35.32878	36.62572	37.54800	38.30599	37.45871	36.94277	37.86217	266.94226	8.61104
CX2	INS	20	34.90223	36.48162	37.44424	38.03219	37.31848	36.66893	37.76419	230.21344	7.42624
CX2	INS	30	**34.59096**	36.49315	37.32896	38.08983	37.28032	36.84764	37.73815	217.89116	7.02875
CX2	INV	10	34.93681	36.59114	37.40389	37.99760	37.31742	36.89372	37.74691	269.12641	8.68150
CX2	INV	20	35.32878	36.64014	37.51341	38.25411	37.47034	36.96007	38.00915	232.88722	7.51249
CX2	INV	30	**34.59096**	36.42398	37.02345	37.68057	37.10931	36.65164	37.57973	220.83009	7.12355

Table A3. Instance 15-2. Statistical summary (runtimes in seconds).

Combin. Method	Improv. Method	Local Search Size	min	Q1	Median	Q2	(Pseudo) Median	lb	ub	Total Runtimes	Mean Runtimes
PMX	EXC	10	**33.34589**	33.37472	33.49576	33.78974	33.60532	33.49580	33.73214	227.75893	7.34706
PMX	EXC	20	**33.34589**	33.47847	33.68022	33.90214	33.72628	33.57073	33.96260	193.70328	6.24849
PMX	EXC	30	**33.34589**	33.37472	33.49000	33.67157	**33.50802**	**33.43241**	**33.72633**	182.34684	5.88216
PMX	INS	10	**33.34589**	33.49288	33.97996	34.43533	33.98722	33.74071	34.24511	263.04966	8.48547
PMX	INS	20	33.46694	33.68886	34.00878	34.39210	34.05206	33.87327	34.21623	229.88504	7.41565
PMX	INS	30	**33.34589**	33.49000	33.52459	34.17018	33.79832	33.56488	33.95395	217.47901	7.01545
PMX	INV	10	**33.34589**	33.43236	33.64563	33.76668	33.66869	33.55622	33.78966	267.02398	8.61368
PMX	INV	20	**33.34589**	33.51017	33.76668	33.95978	33.74652	33.65138	33.87620	230.50383	7.43561
PMX	INV	30	**33.34589**	33.43236	33.55917	33.97707	33.70549	33.52462	33.85884	219.22672	7.07183
OX	EXC	10	**33.34589**	33.44389	33.70328	33.86179	33.71767	33.57075	33.88770	229.91375	7.41657
OX	EXC	20	**33.34589**	33.50729	33.76668	34.38345	33.94542	33.67742	34.10389	195.04656	6.29182
OX	EXC	30	**33.34589**	33.47271	33.64563	33.77245	33.64563	33.56782	33.76668	182.48869	5.88673
OX	INS	10	**33.34589**	33.37472	33.64563	33.81280	33.66285	33.53901	33.76086	264.40133	8.52908
OX	INS	20	**33.34589**	33.37472	33.64563	33.81280	33.66868	33.55627	33.80125	229.29051	7.39647
OX	INS	30	**33.34589**	33.43812	33.63987	33.74939	33.59379	33.51877	33.70334	217.31168	7.01005
OX	INV	10	**33.34589**	33.64275	33.76668	34.32869	33.88645	33.70616	34.06637	268.06627	8.64730
OX	INV	20	**33.34589**	33.64275	33.81280	34.25664	33.91238	33.74359	34.29691	232.63654	7.50440
OX	INV	30	**33.34589**	33.52747	33.76668	34.06930	33.81849	33.66584	34.01446	220.26735	7.10540
CX	EXC	10	**33.34589**	33.37472	33.49576	33.95978	33.66584	33.51021	33.89634	222.08047	7.16389
CX	EXC	20	**33.34589**	33.42083	33.63987	33.91079	33.68013	33.55628	33.83583	189.45881	6.11157
CX	EXC	30	**33.34589**	33.37472	33.70328	34.07507	33.74260	33.57067	34.00014	177.81074	5.73583
CX	INS	10	**33.34589**	33.90502	34.45262	35.04057	34.46412	34.17891	34.76682	259.44712	8.36926
CX	INS	20	33.49000	34.14135	34.59673	35.14144	34.64280	34.35470	34.94826	225.07482	7.26048
CX	INS	30	33.37472	33.54188	34.43533	34.85323	34.30566	34.03188	34.62557	212.64264	6.85944
CX	INV	10	**33.34589**	33.52459	33.65140	34.00013	33.76089	33.61400	33.91940	260.00214	8.38717
CX	INV	20	**33.34589**	33.37472	33.63987	33.89926	33.66286	33.51309	33.88201	227.02476	7.32338
CX	INV	30	**33.34589**	33.62546	33.87620	34.39786	33.97136	33.76384	34.12980	216.41019	6.98097
CX2	EXC	10	**33.34589**	33.53611	34.11830	35.05786	34.35471	33.93387	34.68322	231.42931	7.46546
CX2	EXC	20	33.37472	34.02607	34.57367	34.93681	34.56799	34.28830	34.84753	196.16718	6.32797
CX2	EXC	30	33.37472	33.66004	34.13559	34.86476	34.28402	33.97416	34.69187	183.38571	5.91567
CX2	INS	10	33.49000	34.93393	35.46712	35.78992	35.38606	35.08090	35.68042	265.78683	8.57377
CX2	INS	20	**33.34589**	34.79847	35.46712	35.92826	35.38210	35.03192	35.78415	230.63251	7.43976
CX2	INS	30	33.65140	34.34310	35.34031	35.64293	35.13424	34.79847	35.45847	218.36783	7.04412
CX2	INV	10	33.37472	34.28834	34.71777	35.78415	35.01603	34.58802	35.43836	268.83379	8.67206
CX2	INV	20	**33.34589**	33.79262	34.66013	35.42677	34.72070	34.28255	35.20772	233.39165	7.52876
CX2	INV	30	33.37472	33.66293	34.74660	35.31437	34.72647	34.33444	35.06652	221.32340	7.13946

Table A4. Instance 15-3. Statistical summary (runtimes in seconds).

Combin. Method	Improv. Method	Local Search Size	min	Q1	Median	Q2	(Pseudo) Median	lb	ub	Total Runtimes	Mean Runtimes
PMX	EXC	10	**35.21926**	35.26249	35.58240	35.74957	**35.54210**	**35.45850**	**35.67456**	227.71527	7.34565
PMX	EXC	20	**35.21926**	**35.21926**	35.58240	35.74957	35.54489	35.43258	35.66306	194.81453	6.28434
PMX	EXC	30	**35.21926**	35.29708	35.66310	35.79568	35.55641	35.48434	35.72358	183.07657	5.90570
PMX	INS	10	**35.21926**	35.49306	35.78415	36.02048	35.79617	35.63136	35.94266	266.05218	8.58233
PMX	INS	20	**35.21926**	35.45847	35.71498	35.95996	35.74837	35.58237	35.90234	232.48314	7.49946
PMX	INS	30	**35.21926**	35.58240	35.78415	35.98302	35.78413	35.64009	35.90807	219.23577	7.07212
PMX	INV	10	**35.21926**	**35.21926**	35.62852	35.79568	35.58237	35.46714	35.72358	268.63915	8.66578
PMX	INV	20	**35.21926**	35.37201	35.64581	35.78415	35.62559	35.50167	35.74958	231.71591	7.47471
PMX	INV	30	**35.21926**	**35.21926**	35.40371	35.95420	35.60547	35.44114	35.79568	219.80032	7.09033
OX	EXC	10	**35.21926**	35.31437	35.74957	35.98302	35.74375	35.56508	35.89941	242.16606	7.81181
OX	EXC	20	**35.21926**	35.28843	35.62852	35.88214	35.57376	35.46711	35.76973	194.91102	6.28745
OX	EXC	30	**35.21926**	35.28843	35.57664	35.76109	35.55928	35.44405	35.70923	181.91904	5.86836
OX	INS	10	**35.21926**	35.28843	35.71498	35.79568	35.57664	35.48444	35.74958	264.82807	8.54284
OX	INS	20	**35.21926**	35.28843	35.64581	36.00896	35.64977	35.48437	35.79566	229.24999	7.39516
OX	INS	30	**35.21926**	35.29708	35.68616	35.91961	35.68610	35.50164	35.82444	217.21216	7.00684
OX	INV	10	**35.21926**	**35.21926**	35.66310	36.04066	35.69477	35.47583	35.87061	269.37530	8.68953
OX	INV	20	**35.21926**	35.68904	35.86485	36.17612	35.90231	35.74958	36.09830	233.15601	7.52116
OX	INV	30	**35.21926**	35.57952	35.77262	36.13289	35.80140	35.66313	35.97141	220.73601	7.12052
CX	EXC	10	**35.21926**	35.31437	35.71498	35.88791	35.68137	35.53626	35.80724	223.12399	7.19755
CX	EXC	20	**35.21926**	35.30572	35.74957	36.02048	35.68054	35.53921	35.85622	190.05782	6.13090
CX	EXC	30	**35.21926**	35.35472	35.83603	35.99743	35.79851	35.59677	35.92537	178.46300	5.75687
CX	INS	10	35.66310	36.09254	36.45856	36.89664	36.51188	36.30293	36.73813	258.61783	8.34251
CX	INS	20	**35.21926**	35.73227	36.25682	36.76118	36.24549	36.01479	36.44419	222.54077	7.17873
CX	INS	30	**35.21926**	35.72939	36.06660	36.49891	36.09080	35.90226	36.30580	211.01199	6.80684
CX	INV	10	**35.21926**	35.40371	35.84756	36.50468	35.97941	35.72075	36.19922	259.19600	8.36116
CX	INV	20	**35.21926**	35.74957	36.02625	36.35193	36.03029	35.87642	36.17616	226.44557	7.30470
CX	INV	30	**35.21926**	35.72363	36.15306	36.31734	36.12424	35.93112	36.32315	215.50162	6.95167
CX2	EXC	10	35.28843	35.78415	36.08965	36.75254	36.23662	35.97145	36.47011	231.04708	7.45313
CX2	EXC	20	**35.21926**	35.84468	36.36057	36.60843	36.27787	36.02330	36.49895	194.29054	6.26744
CX2	EXC	30	**35.21926**	35.94843	36.40668	36.78424	36.37367	36.13575	36.64306	182.08140	5.87359
CX2	INS	10	**35.21926**	36.23088	36.75830	37.21079	36.69778	36.38939	37.00328	264.41698	8.52958
CX2	INS	20	35.74957	36.45856	37.00040	37.55952	37.02864	36.72077	37.32900	228.83163	7.38167
CX2	INS	30	35.34031	36.16171	37.00616	37.49612	37.01976	36.59396	37.36073	216.29497	6.97726
CX2	INV	10	**35.21926**	36.28852	36.82747	37.39236	36.82747	36.53062	37.11280	267.69497	8.63532
CX2	INV	20	35.40371	36.49891	36.76983	37.17332	36.81017	36.59400	37.00332	231.97497	7.48306
CX2	INV	30	35.62852	36.30005	36.78136	37.04363	36.70892	36.48733	36.93411	219.82824	7.09123

Table A5. Instance 30-1. Statistical summary (runtimes in seconds).

Combin. Method	Improv. Method	Local Search Size	min	Q1	Median	Q2	(Pseudo) Median	lb	ub	Total Runtimes	Mean Runtimes
PMX	EXC	10	51.06504	53.01911	53.87221	55.39396	54.18930	53.52346	54.83772	881.08183	28.42199
PMX	EXC	20	51.18033	53.42548	54.44863	55.79457	54.54374	53.92409	55.20086	737.33660	23.78505
PMX	EXC	30	51.91238	53.73099	54.49474	55.71675	54.66767	54.02208	55.32479	691.54074	22.30777
PMX	INS	10	52.11989	54.03937	55.56688	56.29606	55.58562	54.85789	56.19518	1037.05853	33.45350
PMX	INS	20	50.59814	54.06531	55.16339	56.55256	55.33776	54.59850	56.14619	894.97129	28.87004
PMX	INS	30	51.80286	52.98452	54.90400	56.01361	54.74001	53.99612	55.47184	846.55571	27.30825
PMX	INV	10	52.45998	54.74837	55.63029	56.65344	55.62599	55.00486	56.22979	1049.03119	33.83972
PMX	INV	20	52.38505	54.78295	56.35658	56.71108	55.94052	55.19506	56.55546	907.41383	29.27141
PMX	INV	30	52.83465	54.71378	55.44007	56.96470	55.59566	55.01924	56.40270	861.09960	27.77741
OX	EXC	10	51.47430	53.61859	54.62732	56.40558	54.83195	54.15177	55.68505	905.71830	29.21672
OX	EXC	20	51.75675	53.43413	54.18348	55.82916	54.48335	53.87801	54.98472	754.31409	24.33271
OX	EXC	30	50.59814	54.25841	55.27291	56.51798	55.40981	54.77719	55.96461	700.88260	22.60912
OX	INS	10	52.56950	53.86068	54.55239	55.61876	54.76566	54.20653	55.36514	1057.72009	34.12000
OX	INS	20	50.60391	52.50321	53.79151	54.93282	**53.78863**	**53.18339**	**54.39675**	902.89111	29.12552
OX	INS	30	50.24653	53.71081	54.22959	55.17780	54.48466	53.95291	55.12304	851.61238	27.47137
OX	INV	10	51.42819	54.16618	55.20374	56.85518	55.38387	54.74837	56.03378	1064.09214	34.32555
OX	INV	20	52.77125	54.39099	56.09719	57.55553	56.22526	55.43138	57.07703	908.57995	29.30903
OX	INV	30	52.29858	54.99911	56.69667	57.31055	56.27156	55.67352	56.98488	859.53252	27.72686
CX	EXC	10	51.66452	53.61570	54.53509	56.41422	54.93138	54.20077	55.50636	871.22559	28.10405
CX	EXC	20	51.57806	53.76557	54.58121	55.61876	54.59850	54.04802	55.12304	725.63137	23.40746
CX	EXC	30	**49.95832**	53.23815	54.28147	54.88671	54.19499	53.60122	54.70222	678.97770	21.90251
CX	INS	10	53.24968	55.71964	57.51518	58.73719	57.18807	56.39117	57.99361	1034.47365	33.37012
CX	INS	20	52.91535	55.81475	56.93300	57.86392	56.76440	56.18077	57.33073	882.38903	28.46416
CX	INS	30	53.11134	55.20086	56.17789	57.46907	56.22405	55.56981	56.95313	832.18669	26.84473
CX	INV	10	53.23815	55.70811	57.17510	58.92453	57.26012	56.43151	57.99938	1031.70938	33.28095
CX	INV	20	53.77422	55.40837	56.99064	57.98208	56.79170	56.10001	57.48638	893.67188	28.82813
CX	INV	30	53.61282	55.86662	56.58715	57.94462	56.78169	56.25282	57.35955	842.95357	27.19205
CX2	EXC	10	54.06819	55.67352	56.84654	58.18960	56.98344	56.27012	57.71981	936.78281	30.21880
CX2	EXC	20	52.18330	54.98182	56.52951	58.01955	56.34582	55.46310	57.23273	768.74164	24.79812
CX2	EXC	30	51.43972	55.82916	56.77160	57.58724	56.66455	55.92714	57.22404	711.29197	22.94490
CX2	INS	10	53.67623	55.57841	57.40566	59.18680	57.41143	56.59291	58.46916	1092.76569	35.25051
CX2	INS	20	54.44287	56.83501	58.17230	59.30209	58.09449	57.32208	58.91300	921.59958	29.72902
CX2	INS	30	54.12007	57.07422	58.22418	59.97073	58.26309	57.54977	59.12628	862.70223	27.82910
CX2	INV	10	54.64461	57.81780	58.97929	60.37134	59.11907	58.35099	59.77475	1097.34856	35.39834
CX2	INV	20	55.37090	57.79475	58.73143	60.02549	58.88416	58.22412	59.56438	931.48914	30.04804
CX2	INV	30	55.55536	57.01370	59.48654	60.08313	58.82654	57.99361	59.69982	873.67884	28.18319

Table A6. Instance 30-2. Statistical summary (runtimes in seconds).

Combin. Method	Improv. Method	Local Search Size	min	Q1	Median	Q2	(Pseudo) Median	lb	ub	Total Runtimes	Mean Runtimes
PMX	EXC	10	49.74504	50.53474	51.48583	52.05937	51.35325	50.99587	51.75098	866.66081	27.95680
PMX	EXC	20	49.94679	50.46845	51.42819	52.03919	**51.25815**	**50.92378**	**51.57232**	722.85330	23.31785
PMX	EXC	30	49.08792	50.93535	51.83168	52.38793	51.71461	51.29557	52.11989	675.78662	21.79957
PMX	INS	10	50.06207	51.80286	52.39657	52.93841	52.38352	51.93826	52.82593	1017.67223	32.82814
PMX	INS	20	50.73649	51.46566	52.61561	53.16321	52.45854	52.03055	52.92112	876.12927	28.26223
PMX	INS	30	50.06207	51.92103	52.83465	53.33614	52.76837	52.19771	53.14304	827.92987	26.70742
PMX	INV	10	49.89491	51.62994	52.33317	52.94706	52.26113	51.81724	52.71078	1028.13709	33.16571
PMX	INV	20	49.83727	51.60976	52.59256	53.28138	52.50612	52.08530	52.94996	886.08724	28.58346
PMX	INV	30	51.27832	52.52627	53.02487	53.73387	53.17474	52.71361	53.61570	837.62361	27.02012
OX	EXC	10	49.37037	50.88635	51.65876	52.55221	51.67893	51.26391	52.10836	895.56372	28.88915
OX	EXC	20	49.58941	50.75954	51.50312	52.30435	51.55788	51.14574	52.01902	740.58223	23.88975
OX	EXC	30	49.01299	50.79989	51.38208	52.10836	51.49444	51.06503	51.97289	683.61961	22.05225
OX	INS	10	49.02452	50.57221	51.41666	51.94985	51.35324	50.95548	51.73083	1038.97999	33.51548
OX	INS	20	49.88338	50.86618	51.64723	52.45710	51.71784	51.26103	52.12566	885.74781	28.57251
OX	INS	30	**48.69596**	50.71919	51.44548	52.63579	51.64291	51.09963	52.13142	834.56697	26.92152
OX	INV	10	50.48286	52.20636	53.24968	54.12295	53.29435	52.62714	53.97597	1062.18632	34.26407
OX	INV	20	50.67308	52.33029	52.69055	53.60417	52.87072	52.53198	53.26700	901.22178	29.07167
OX	INV	30	51.19186	52.32164	53.26120	54.10278	53.19059	52.66749	53.75693	840.90345	27.12592
CX	EXC	10	50.41945	51.11692	51.92967	52.49457	51.88065	51.52046	52.26403	862.95733	27.83733
CX	EXC	20	49.67587	51.18321	51.57806	52.00173	51.57229	51.26679	51.86627	714.57964	23.05096
CX	EXC	30	49.16286	51.17745	51.53771	52.20924	51.63646	51.30712	52.05644	667.59255	21.53524
CX	INS	10	52.06225	53.46872	54.12007	55.50060	54.42113	53.91260	55.06257	1018.25457	32.84692
CX	INS	20	49.43378	52.99605	54.24688	55.59282	54.28946	53.56378	54.96741	870.52058	28.08131
CX	INS	30	51.26679	53.56959	54.81754	55.93291	54.81466	54.18348	55.38243	818.01102	26.38745
CX	INV	10	51.26679	53.46007	54.40252	55.13169	54.28145	53.79733	54.85797	1031.82092	33.28455
CX	INV	20	50.55203	52.97299	54.14889	55.07116	53.98365	53.41391	54.56969	878.18085	28.32841
CX	INV	30	51.72216	53.48024	54.45439	54.87518	54.18346	53.72225	54.63600	829.36701	26.75377
CX2	EXC	10	50.03902	52.26400	53.00758	53.74828	53.02915	52.52047	53.52927	926.92980	29.90096
CX2	EXC	20	50.64426	52.53492	53.39378	54.41693	53.42838	52.90093	53.94423	753.77314	24.31526
CX2	EXC	30	49.29544	52.77989	53.46295	54.33335	53.53701	53.03640	54.04514	694.61395	22.40690
CX2	INS	10	53.13439	54.16042	56.07990	57.96191	56.13610	55.41990	57.08287	1075.41473	34.69080
CX2	INS	20	54.31029	55.98191	57.20392	58.14060	57.16501	56.58138	57.78034	901.75945	29.08901
CX2	INS	30	52.19483	54.17771	56.18942	57.41719	56.06261	55.23833	56.99064	844.93842	27.25608
CX2	INV	10	53.14016	55.37955	57.09440	59.13492	57.36532	56.37963	58.25307	1078.62075	34.79422
CX2	INV	20	52.62138	54.54662	55.90121	57.41719	55.97614	55.17492	56.76584	908.90025	29.31936
CX2	INV	30	52.78278	55.54959	56.89841	58.22418	56.89121	56.17213	57.59300	852.43323	27.49785

Table A7. Instance 30-3. Statistical summary (runtimes in seconds).

Combin. Method	Improv. Method	Local Search Size	min	Q1	Median	Q2	(Pseudo) Median	lb	ub	Total Runtimes	Mean Runtimes
PMX	EXC	10	48.80548	50.31282	51.12269	52.18906	51.19762	50.66731	51.71063	878.12475	28.32660
PMX	EXC	20	48.86888	50.24653	51.19762	51.78557	51.08802	50.69614	51.61267	737.05760	23.77605
PMX	EXC	30	49.00146	50.82871	51.43972	52.25535	51.50893	51.06798	51.91244	687.98596	22.19310
PMX	INS	10	49.34155	51.71928	52.65020	53.71081	52.75107	51.96426	53.35920	1034.64337	33.37559
PMX	INS	20	49.98138	50.97858	52.12566	53.48024	52.35316	51.73365	52.97304	889.41709	28.69087
PMX	INS	30	50.32146	52.02190	52.27553	53.74540	52.69156	52.18037	53.16895	839.41356	27.07786
PMX	INV	10	50.75378	51.88068	52.77701	53.37937	52.60985	52.18330	53.03928	1044.30079	33.68712
PMX	INV	20	50.32146	52.21212	52.89230	53.93562	52.99758	52.47434	53.56091	898.29205	28.97716
PMX	INV	30	49.53177	51.54924	52.67902	53.77422	52.72513	52.09107	53.35055	849.39456	27.39982
OX	EXC	10	49.30696	50.99876	51.77981	52.55509	51.83583	51.31862	52.39653	902.29349	29.10624
OX	EXC	20	**48.04460**	50.21194	50.92094	51.39360	**50.84456**	**50.44251**	**51.18033**	746.78063	24.08970
OX	EXC	30	49.35884	50.72784	51.75675	52.20636	51.53190	51.08230	51.93259	693.01672	22.35538
OX	INS	10	48.37316	50.33587	50.93247	51.41378	50.88780	50.50015	51.25238	1047.95678	33.80506
OX	INS	20	48.67290	49.97561	50.95552	52.28129	51.12845	50.58085	51.65011	898.56316	28.98591
OX	INS	30	48.41928	50.44251	51.58959	52.44845	51.47142	50.92670	52.00749	843.49820	27.20962
OX	INV	10	50.23500	51.75387	53.35920	54.48898	53.15313	52.50898	53.77134	1059.09986	34.16451
OX	INV	20	50.18312	51.58670	52.55797	54.14025	52.80439	52.11989	53.38225	902.16492	29.10209
OX	INV	30	49.44530	53.01046	53.50907	54.46016	53.54265	53.01046	54.09119	850.08570	27.42212
CX	EXC	10	49.48565	50.92382	51.81439	52.36487	51.75675	51.25526	52.14295	868.52623	28.01698
CX	EXC	20	48.88041	50.47421	51.70487	52.40234	51.55848	51.08527	52.07383	723.63057	23.34292
CX	EXC	30	49.69893	50.71055	51.46854	52.58679	51.56590	51.08810	52.04494	674.05238	21.74363
CX	INS	10	51.36478	53.25832	54.80025	55.67640	54.65565	53.92409	55.41418	1034.39624	33.36762
CX	INS	20	50.30417	52.67326	54.29876	55.59859	54.42990	53.52059	55.25562	876.13563	28.26244
CX	INS	30	51.71063	53.73387	55.01352	55.96750	54.90544	54.22383	55.49483	825.90955	26.64224
CX	INV	10	51.52042	53.36208	54.66767	55.97614	54.69505	54.03649	55.37667	1040.66017	33.56968
CX	INV	20	52.21788	54.01920	54.92706	55.78592	55.00488	54.51204	55.53518	885.35162	28.55973
CX	INV	30	50.79989	53.96155	54.99046	55.87527	54.94723	54.32758	55.52077	836.78412	26.99304
CX2	EXC	10	50.10242	52.08531	53.34190	54.89536	53.43557	52.66749	54.18924	937.06006	30.22774
CX2	EXC	20	50.20041	52.57238	53.72234	54.74549	53.70649	53.03064	54.48321	764.68281	24.66719
CX2	EXC	30	49.74504	51.88933	53.54365	54.41981	53.26409	52.64732	54.01055	705.87219	22.77007
CX2	INS	10	50.43675	55.15763	56.20095	57.53824	56.22256	55.56400	56.93588	1071.65137	34.56940
CX2	INS	20	51.17456	54.32470	56.20671	57.95038	56.17357	55.27579	57.12034	904.03375	29.16238
CX2	INS	30	52.90959	55.56688	57.45754	59.60759	57.51518	56.47475	58.59597	845.40639	27.27117
CX2	INV	10	53.34767	56.06549	57.84951	59.19833	57.70828	56.74855	58.57580	1076.21986	34.71677
CX2	INV	20	54.40828	56.43440	57.59588	58.92165	57.78786	56.99640	58.69108	910.16465	29.36015
CX2	INV	30	52.94417	55.71099	56.89265	58.72278	56.89121	56.17213	57.59300	853.29243	27.52556

References

1. Szpilko, D.; De la Torre, A.; Jimenez Naharro, F.; Rzepka, A.; Remiszewska, A. Waste management in the smart city: Current practices and future directions. *Resources* **2023**, *12*, 115. [CrossRef]
2. Hess, C.; Dragomir, A.G.; Doerner, K.F.; Vigo, D. Waste collection routing: A survey on problems and methods. *Cent. Eur. J. Oper. Res.* **2024**, *32*, 399–434. [CrossRef]
3. Ghiani, G.; Laganà, D.; Manni, E.; Musmanno, R.; Vigo, D. Operations research in solid waste management: A survey of strategic and tactical issues. *Comput. Oper. Res.* **2014**, *44*, 22–32. [CrossRef]
4. Yaman, C.; Anil, I.; Jaunich, M.; Blaisi, N.; Alagha, O.; Yaman, A.; Gunday, S. Investigation and modelling of greenhouse gas emissions resulting from waste collection and transport activities. *Waste Manag. Res.* **2019**, *37*, 1282–1290. [CrossRef]
5. Rossit, D.; Nesmachnow, S.; Cavallin, A. Municipal Solid waste management systems: Application of SWOT methodology to analyze an Argentinean case study. In Proceedings of the VI Ibero-American Congress of Smart Cities, Mexico City, Mexico, 13–17 November 2023; pp. 668–680.
6. Bonomo, F.; Durán, G.; Larumbe, F.; Marenco, J. A method for optimizing waste collection using mathematical programming: A Buenos Aires case study. *Waste Manag. Res.* **2012**, *30*, 311–324. [CrossRef]
7. Braier, G.; Durán, G.; Marenco, J.; Wesner, F. An integer programming approach to a real-world recyclable waste collection problem in Argentina. *Waste Manag. Res.* **2017**, *35*, 525–533. [CrossRef]
8. Molfese, S.; Rossit, D.; Frutos, M.; Cavallin, A. Optimization of waste collection through the sequencing of micro-routes and transfer station convenience analysis: An Argentinian case study. *Waste Manag. Res.* **2023**, *41*, 1267–1279. [CrossRef]
9. Rossit, D.; Toutouh, J.; Nesmachnow, S. Exact and heuristic approaches for multi-objective garbage accumulation points location in real scenarios. *Waste Manag.* **2020**, *105*, 467–481. [CrossRef]
10. Cavallin, A.; Rossit, D.; Herran, V.; Rossit, D.; Frutos, M. Application of a methodology to design a municipal waste pre-collection network in real scenarios. *Waste Manag. Res.* **2020**, *38*, 117–129. [CrossRef]
11. Rossit, D.; González Landín, B.; Frutos, M.; Méndez Babey, M. An Allocation-Routing Problem in Waste Management Planning: Exact and Heuristic Resolution Approaches. In *Ibero-American Congress of Smart Cities*; Nesmachnow, S., Hernández, L., Eds.; Springer: Cham, Switzerland, 2023; pp. 92–107. [CrossRef]
12. Hannan, M.; Lipu, M.; Akhtar, M.; Begum, R.; Al Mamun, M.; Hussain, A.; Mia, M.; Basri, H. Solid waste collection optimization objectives, constraints, modeling approaches, and their challenges toward achieving sustainable development goals. *J. Clean. Prod.* **2020**, *277*, 123557. [CrossRef]
13. Toth, P.; Vigo, D. *Vehicle Routing: Problems, Methods, and Applications*; SIAM: Philadelphia, PA, USA, 2014. [CrossRef]
14. Elshaer, R.; Awad, H. A taxonomic review of metaheuristic algorithms for solving the vehicle routing problem and its variants. *Comput. Ind. Eng.* **2020**, *140*, 106242. [CrossRef]
15. Kalra, M.; Tyagi, S.; Kumar, V.; Kaur, M.; Mashwani, W.; Shah, H.; Shah, K. A comprehensive review on scatter search: Techniques, applications, and challenges. *Math. Probl. Eng.* **2021**, *2021*, 5588486. [CrossRef]
16. Liang, Y.; Minanda, V.; Gunawan, A. Waste collection routing problem: A mini-review of recent heuristic approaches and applications. *Waste Manag. Res.* **2022**, *40*, 519–537. [CrossRef] [PubMed]
17. Zhang, Q.; Wu, L.; Li, J. Application of Improved Scatter Search Algorithm to Reverse Logistics VRP Problem. In Proceedings of the 2023 IEEE 5th International Conference on Power, Intelligent Computing and Systems (ICPICS), Shenyang, China, 14–16 July 2023; pp. 283–288. [CrossRef]
18. Dethloff, J. Vehicle routing and reverse logistics: The vehicle routing problem with simultaneous delivery and pick-up: Fahrzeugeinsatzplanung und Redistribution: Tourenplanung mit simultaner Auslieferung und Rückholung. *OR-Spektrum* **2001**, *23*, 79–96. [CrossRef]
19. Chu, F.; Labadi, N.; Prins, C. A scatter search for the periodic capacitated arc routing problem. *Eur. J. Oper. Res.* **2006**, *169*, 586–605. [CrossRef]
20. Golden, B.L.; DeArmon, J.S.; Baker, E.K. Computational experiments with algorithms for a class of routing problems. *Comput. Oper. Res.* **1983**, *10*, 47–59. [CrossRef]
21. Belenguer, J.M.; Benavent, E. A cutting plane algorithm for the capacitated arc routing problem. *Comput. Oper. Res.* **2003**, *30*, 705–728. [CrossRef]
22. Yu, H.; Li, X.; Su, S. Material recovery facility location and waste flow assignment problem based on scatter search. In Proceedings of the 2007 International Conference on Machine Learning and Cybernetics, Hong Kong, China, 19–22 August 2007; Volume 4, pp. 2368–2372. [CrossRef]
23. Ortega, M.; Hipólito, J.D.; García, Á. Scatter search for locating a treatment plant and the necessary transfer centers in a reverse network. In *Metaheuristics in the Service Industry*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 63–81. [CrossRef]
24. Mahéo, A.; Rossit, D.; Kilby, P. Solving the integrated bin allocation and collection routing problem for municipal solid waste: A Benders decomposition approach. *Ann. Oper. Res.* **2023**, *322*, 441–465. [CrossRef]
25. Dror, M.; Laporte, G.; Trudeau, P. Vehicle routing with split deliveries. *Discret. Appl. Math.* **1994**, *50*, 239–254. [CrossRef]
26. NOES Server Solvers. Available online: https://neos-server.org/neos/solvers/index.html (accessed on 27 October 2024).
27. Linderoth, J.; Ralphs, T. Noncommercial software for mixed-integer linear programming. In *Integer Programming*; CRC Press: London, UK, 2005; pp. 269–320. [CrossRef]

28. Meindl, B.; Templ, M. *Analysis of Commercial and Free and Open Source Solvers for Linear Optimization Problems*; Technical Report CS-2012-1; Vienna University of Technology: Vienna, Austria, 2012. Available online: http://hdl.handle.net/20.500.12708/37465 (accessed on 27 October 2024).
29. Atamtürk, A.; Savelsbergh, M.W. Integer-programming software systems. *Ann. Oper. Res.* **2005**, *140*, 67–124. [CrossRef]
30. Du, Y.; Kochenberger, G.; Glover, F.; Wang, H.; Lewis, M.; Xie, W.; Tsuyuguchi, T. Solving clique partitioning problems: A comparison of models and commercial solvers. *Int. J. Inf. Technol. Decis. Mak.* **2022**, *21*, 59–81. [CrossRef]
31. González, M.; López, J.; Aparicio, J. A parallel algorithm for matheuristics: A comparison of optimization solvers. *Electronics* **2020**, *9*, 1541. [CrossRef]
32. Gurobi Optimization, LLC. Gurobi Optimizer Reference Manual. 2024. Available online: https://www.gurobi.com (accessed on 27 October 2024).
33. International Business Machines Corporation. IBM ILOG CPLEX Optimization Studio Reference Manual. 2024. Available online: https://www.ibm.com/docs/en/icos/22.1.1 (accessed on 27 October 2024).
34. Hart, W.; Watson, J.; Woodruff, D. Pyomo: Modeling and solving mathematical programs in Python. *Math. Program. Comput.* **2011**, *3*, 219–260. [CrossRef]
35. Glover, F. Heuristics for integer programming using surrogate constraints. *Decis. Sci.* **1977**, *8*, 156–166. [CrossRef]
36. Martí, R.; Laguna, M.; Glover, F. Principles of scatter search. *Eur. J. Oper. Res.* **2006**, *169*, 359–372. [CrossRef]
37. Laguna, M.; Martí, R. Scatter Search. In *Metaheuristic Procedures for Training Neural Networks*; Operations Research/Computer Science Interfaces Series; Alba, E., Martí, R., Eds.; Springer: Boston, MA, USA, 2006; Volume 36, pp. 139–152. [CrossRef]
38. Carlos, M.; Gallardo, A.; Edo-Alcon, N.; Abaso, J.R. Influence of the municipal solid waste collection system on the time spent at a collection point: A case study. *Sustainability* **2019**, *11*, 6481. [CrossRef]
39. Owusu, F.; Oduro, S.; Essandoh, H.; Wayo, F.; Shamudeen, M. Characteristics and management of landfill solid waste in Kumasi, Ghana. *Sci. Afr.* **2019**, *3*, e00052. 0 052 [CrossRef]
40. D'Onza, G.; Greco, G.; Allegrini, M. Full cost accounting in the analysis of separated waste collection efficiency: A methodological proposal. *J. Environ. Manag.* **2016**, *167*, 59–65. [CrossRef]
41. Pichugina, O.; Matsiy, O.; Skob, Y. Performance Comparison of Unbounded Knapsack Problem Formulations. *CEUR-WS* **2023**, *3403*, 263–272.
42. Saghand, P.G.; Charkhgard, H. Exact solution approaches for integer linear generalized maximum multiplicative programs through the lens of multi-objective optimization. *Comput. Oper. Res.* **2022**, *137*, 105549. [CrossRef]
43. Schulze, P.; Walter, R. A note on the exact solution of the minimum squared load assignment problem. *Comput. Oper. Res.* **2023**, *159*, 106309. [CrossRef]
44. Goldberg, D.; Lingle, R. Alleles, Loci and the traveling salesman problem. In *Proceedings of the First International Conference on Genetic Algorithms and Their Applications*; Psychology Press: New York, NY, USA, 2014; pp. 154–159. [CrossRef]
45. Davis, L. Applying adaptive algorithms to epistatic domains. In Proceedings of the International Joint Conferences on Artificial Intelligence, San Francisco, CA, USA, 18–23 August 1985; pp. 162–164. [CrossRef]
46. Oliver, I.; Smith, D.; Holland, J. A study of permutation crossover operators on the traveling salesman problem. In *Proceedings of the Second International Conference on Genetic Algorithms on Genetic Algorithms and Their Application*; Psychology Press: New York, NY, USA, 1987; pp. 224–230. [CrossRef]
47. Hussain, A.; Muhammad, Y.; Nauman, M.; Hussain, I.; Mohamd, A.; Gani, S. Genetic algorithm for traveling salesman problem with modified cycle crossover operator. *Comput. Intell. Neurosci.* **2017**, *1*, 7430125. [CrossRef]
48. Hollander, M.; Wolfe, D. *Nonparametric Statistical Methods*; John Wiley & Sons: Hoboken, NJ, USA, 1973.
49. Siegel, S. *Non Parametric Statistics for the Behavioural Sciences*; MacGraw Hill Int: Columbus, OH, USA, 1988.
50. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022. Available online: https://www.R-project.org/ (accessed on 27 October 2024).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

sustainability

MDPI

Article

Integrating Technology and Urban Resilience: A Comprehensive Analysis of Smart City Initiatives in Sydney

Shabnam Varzeshi, John Fien * and Leila Irajifar

School of Architecture & Urban Design, RMIT University, Melbourne 3000, Australia;
s3941055@student.rmit.edu.au (S.V.); leila.irajifar@rmit.edu.au (L.I.)
* Correspondence: john.fien@rmit.edu.au; Tel.: +61-(0)417-747-100

Abstract: This study explores the integration of smart city and resilience strategies in Sydney, focusing on the relationship between technological advancements and urban resilience. By analysing strategic documents and key projects—specifically the NSW Spatial Digital Twin, Land iQ, and SIMPaCT—this research identifies important synergies and gaps in Sydney's urban planning efforts. The findings indicate that these projects improve urban functionality through real-time data integration, predictive planning, and adaptive infrastructure. However, there are inconsistencies between strategic objectives and actual implementation, particularly concerning stakeholder inclusivity and equity considerations. The study concludes that utilising technologies such as the Internet of Things (IoT) and artificial intelligence (AI), along with equitable stakeholder engagement, has the potential to significantly enhance Sydney's ability to address environmental, social, and economic challenges. These insights offer practical recommendations for policymakers and urban planners who aim to balance innovation with inclusivity in smart city development.

Keywords: data-driven urban development; Digital Twins; environmental sustainability; smart cities; stakeholder engagement; urban resilience

1. Introduction

The 21st century presents unprecedented urban challenges driven by climate change, rapid urbanisation, and technological disruptions. These factors significantly impact urban infrastructures, economies, and communities, necessitating innovative approaches to urban planning and resilience. Climate change causes extreme weather events and environmental stresses that present significant risks to urban systems and residents' well-being [1]. Simultaneously, rapid urbanisation intensifies existing strains, leading to congestion, pollution, inadequate housing, and inadequate services. While advanced technologies provide potential solutions, they also complicate managing privacy, security, and inequality [2]. Addressing these multifaceted challenges requires integrated strategies that enhance urban sustainability and resilience.

In response, smart cities have emerged as a critical strategy to enhance urban governance, infrastructure, and quality of life [3]. By integrating technologies such as the Internet of Things (IoT), artificial intelligence (AI), and big data analytics, smart cities aim to address environmental, social, and economic challenges holistically. Cities worldwide—from Asia to the Americas—are adopting digital solutions to innovate governance and infrastructure, thereby enhancing residents' lives [4,5].

This paper explores how smart city initiatives contribute to enhancing urban resilience through a case study of Sydney, Australia's most populous urban centre. With a metropolitan population of over five million people, Sydney faces significant urban pressures that require innovative and resilient solutions. Recognised as a leading global city, Sydney ranks 18th out of 109 cities in the 2020 Smart City Index, published by the Institute for Management Development (IMD) and Singapore University for Technology and Design

(SUTD). This ranking reflects the positive perception of Sydney's citizens regarding the impact of technology on their quality of life [6]. However, the city faces challenges such as managing population growth, mitigating environmental pressures, and enhancing sustainable transportation, compounded by vulnerabilities to climate change impacts such as rising sea levels [7,8].

To navigate these complexities, Sydney has implemented comprehensive resilience strategies and has participated in the Rockefeller Foundation's 100 Resilient Cities Initiative, which is designed to help cities better withstand and recover from adversities [9]. Urban resilience, as defined in this study, refers to a city's ability to absorb, adapt to, and recover from shocks and stressors while maintaining essential functions and promoting sustainable development [10]. To achieve meaningful and deep sustainability, a city must address underlying problems such as social inequality, economic instability, and climate change, which exacerbate the impacts of disruptions like environmental hazards and social distress [11]. This comprehensive understanding of resilience aligns with both the *Resilient Sydney Strategy* and the *Smart City Strategic Framework*, as Sydney seeks to tackle environmental pressures, social inequalities, and economic challenges through integrated smart city initiatives. Guided by these strategies, Sydney utilises IoT, AI, and big data analytics while emphasising strong stakeholder engagement in disaster risk reduction and environmental sustainability [12,13].

By examining key projects such as the NSW Spatial Digital Twin, Land iQ, and SIMPaCT within these frameworks, this study analyses the impact of advanced technologies in fostering resilient urban planning. This provides valuable insights into the practical implementation of smart city strategies, offering lessons applicable to Australia and other urban regions worldwide.

1.1. Research Gap

Despite advancements in global smart city initiatives, significant gaps still exist in understanding how specific technologies and strategies contribute to urban resilience, particularly in Australia. There has been limited examination of the actual impacts of smart city projects, often revealing a disconnect between strategic planning and implementation. Furthermore, the combined effects of technological advancements and resilience initiatives remain underexplored, especially when it comes to social factors that can weaken resilience.

1.2. Research Questions

To address these gaps, this study focuses on the following research questions:

1. How do Sydney's Smart City Strategic Framework and Resilient Sydney Strategy employ advanced technologies such as IoT, AI, and big data to improve urban resilience?
2. What are the synergies and discrepancies between these strategic frameworks, and how do they affect the implementation of smart city initiatives in Sydney?
3. How do specific smart city projects—namely, the NSW Spatial Digital Twin, Land iQ, and SIMPaCT—enhance urban functionality, promote environmental sustainability, and encourage community inclusion in Sydney?
4. What challenges and limitations are encountered in implementing smart city initiatives in Sydney, and what strategies can be adopted to address these issues to ensure long-term resilience?

By addressing these questions, the study aims to provide a comprehensive analysis of how technological advancements can help address the social factors that undermine resilience. The methods used include a two-phased approach combining descriptive, content, comparative, critical, and typological analyses. The following sections evaluate how these projects contribute to developing smarter and more resilient urban environments to draw meaningful conclusions for urban planners and policymakers worldwide.

2. Literature Review

Smart city technologies and processes have evolved significantly in developing and managing cities in many parts of the world. The primary objective of smart city initiatives was to increase urban efficiency by integrating Information and Communication Technology (ICT). With increasingly complex challenges facing cities worldwide, smart city technologies have expanded in scope to include IoT, AI, big data analytics, and more advanced frameworks like Geospatial Artificial Intelligence (GeoAI) [14]. These innovations are vital for improving urban management and decision-making. They enable real-time monitoring and analysis of urban dynamics, allowing cities to enhance the responsiveness and efficiency of services. The application of these technologies has also expanded to include support for disaster risk reduction and management, thereby helping to build urban resilience [8,15]. For example, deploying IoT sensors in urban infrastructure enables real-time data collection on traffic, air quality, and energy consumption, which can be analysed using AI algorithms to optimise city operations [16].

Smart city technologies are increasingly recognised as essential tools for tackling complex urban challenges, including environmental sustainability, social equity, and urban resilience [17,18]. Advanced tools like AI and big data analytics enhance urban governance by enabling data-driven decision-making processes [19]. Integrating ICT with environmental and social considerations is crucial for creating sustainable urban environments, ensuring that technological advancements align with broader societal goals [20]. These technologies offer a framework for cities to manage resources efficiently, address inequalities, and foster resilience in the face of various challenges.

2.1. Urban Resilience in Smart City Frameworks

Urban resilience refers to a city's capacity to withstand, recover from, and adapt to various challenges, such as climate change, economic fluctuations, and social inequalities [10]. As cities encounter more frequent and severe disruptions, integrating resilience into smart city frameworks has become increasingly important [21]. Advanced technologies, including predictive analytics and comprehensive sensor networks, play a crucial role in this process. For instance, cities like Amsterdam and Singapore have implemented smart infrastructure systems that facilitate proactive urban management, effectively mitigating stressors to prevent escalation into significant issues [10,22].

Research shows that incorporating resilience into smart city planning improves both the physical durability of urban infrastructure and the adaptive capabilities of social and economic systems [23]. Resilient smart cities utilise technology to anticipate potential threats, strategise effectively, and recover quickly from adverse events. This approach helps ensure the continuity of services and protects the well-being of citizens [24].

The integration of resilience into smart city frameworks directly contributes to achieving global sustainability objectives. Specifically, the Sustainable Development Goals (SDGs) serve as a guide for tackling critical urban challenges, as outlined below.

2.2. Smart Cities and the SDGs

The United Nations' Sustainable Development Goals (SDGs) offer a global framework for tackling pressing issues like poverty, inequality, climate change, and environmental degradation [25]. Smart city initiatives are strongly connected to several Sustainable Development Goals (SDGs), especially those concerning sustainable cities and communities (SDG 11), affordable and clean energy (SDG 7), and climate action (SDG 13) [26]. By utilising advanced technologies, smart cities can help achieve these goals by managing resources efficiently, minimising environmental impacts, and improving residents' quality of life [27].

The use of renewable energy solutions and smart grids in cities such as Copenhagen and Vancouver has significantly lowered greenhouse gas emissions, thereby supporting SDG 13 [28]. Smart mobility solutions, such as intelligent transportation systems and shared mobility services, enhance sustainable urban transport, aligning with SDG 11 [29].

2.3. Promoting Social Equity in Smart Cities Through Technology

However, questions have been asked about whether smart technologies are capable of addressing the social roots of vulnerability, such as social exclusion and economic disparity. This is because inequalities may worsen in a smart city if access to technology is unequal [20]. Therefore, technological advancements must be integrated with social equity frameworks to ensure inclusive resilience. Strengthening urban resilience for all demands, fostering economic vitality, improving communication, and facilitating citizen participation—and smart technologies can contribute to these three goals [30–32]. For example, Copenhagen's renewable energy initiatives promote environmental sustainability and social equity by ensuring all citizens can access clean energy. Similarly, the smart city strategies of Barcelona emphasise public participation in urban planning, aligning technological advancements with the needs and priorities of the public [33]. These approaches support SDG 10 by fostering inclusivity and empowering marginalised communities. Many smart city projects, however, struggle to align technological progress with social goals, raising concerns about digital divides and the potential to leave certain populations behind [34].

2.4. Gaps in Existing Research

Despite an increasing amount of literature on smart cities, significant gaps still exist, especially regarding the effects of smart city initiatives on urban resilience and sustainability. Much of the existing research emphasises technological innovations or strategic frameworks, but it often fails to thoroughly assess their practical implementation and outcomes [10,22]. In Australia, there is limited research on how smart city initiatives enhance resilience and sustainability.

Additionally, a more comprehensive analysis is needed to examine the combined effects of technological advancements, social equity, and environmental sustainability [35]. This involves exploring how smart technologies can tackle social issues that compromise resilience, such as poverty and social exclusion [36]. Moreover, aligning smart city initiatives with the SDGs offers an opportunity to enhance global sustainability efforts; however, this connection is often insufficiently explored in current studies [37].

This study aims to address existing gaps by examining key smart city projects in Sydney, including the NSW Spatial Digital Twin, Land iQ, and SIMPaCT. It analyses the impact of these initiatives on urban resilience and sustainability. By exploring how these projects integrate advanced technologies with social and environmental objectives, the study enhances our understanding of smart city development in Australia and its alignment with the SDGs.

3. Research Design

This study uses a two-phased approach to analyse Sydney's smart city and resilience strategies. Figure 1 illustrates the research design, highlighting how each analytical method contributes to specific outcomes. In Phase 1, we analyse strategic documents using descriptive, content, comparative, and critical analyses. Phase 2 involves a typological analysis and a detailed examination of three key projects to understand their contributions to urban resilience. These methods were selected to capture the strategic intent behind Sydney's smart city initiatives and the practical implementation of specific projects, ensuring a comprehensive understanding of policy impact, stakeholder engagement, and project effectiveness.

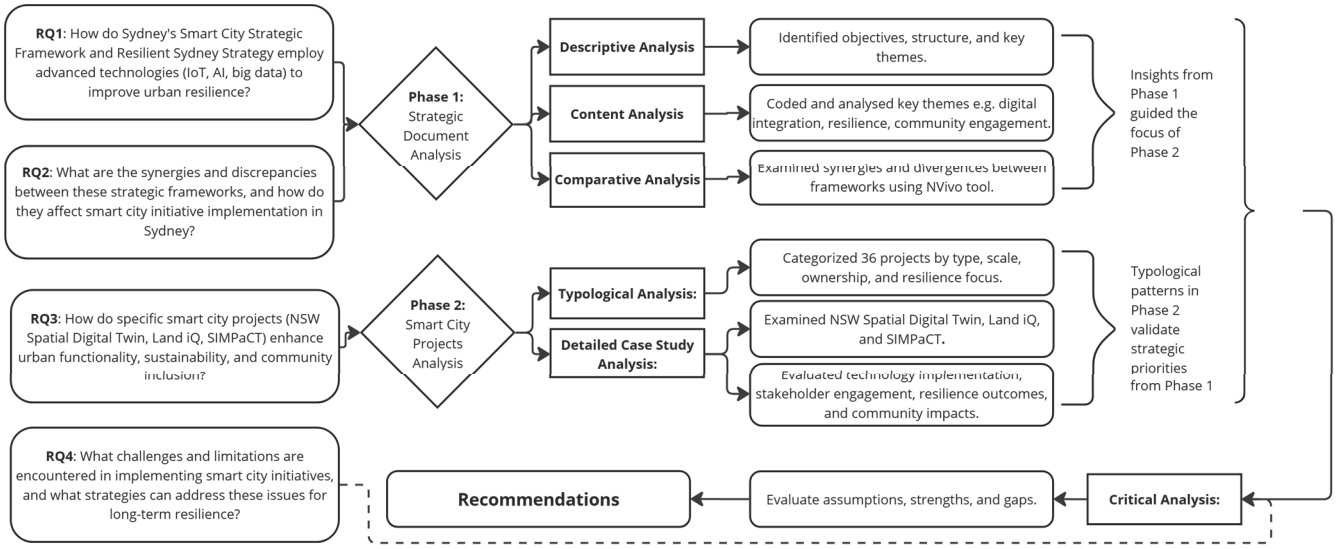


Figure 1. Research design overview.

3.1. Case Study Context

Sydney is an insightful case study that highlights the integration of global smart city trends with local socio-economic and environmental priorities. The city effectively utilises technologies such as AI, IoT, and big data analytics to tackle significant urban challenges related to infrastructure, sustainability, and inclusivity. Key initiatives, like the NSW Spatial Digital Twin and Land iQ, showcase Sydney's ability to monitor conditions in real time and engage in scenario planning, thereby enhancing urban resilience. Additionally, the SIMPaCT initiative aims to optimise water usage and strengthen climate resilience by improving the management of public spaces during extreme weather events. These combined efforts make Sydney an ideal subject for examining the interaction between technology, governance, and resilience.

3.2. Phase 1: Analysis of Strategic Documents

The first phase examines the two key strategic documents: Sydney's *Smart City Strategic Framework* [38] and the *Resilient Sydney Strategy* [39], which guides the city's smart city and resilience efforts. These documents were sourced directly from the official websites of the City of Sydney and Resilient Sydney to ensure their authenticity and accuracy. They were selected because they provide the most comprehensive and up-to-date insights into Sydney's strategic approaches to smart city development and urban resilience, encompassing initiatives from their publication dates through to 2023.

The first step was to conduct a Descriptive Analysis to identify each document's structure, content and key themes. The process involved reading the documents thoroughly to determine the main objectives, strategic goals, and proposed initiatives to lay the groundwork for further analysis [40,41].

Second, Content Analysis techniques developed by Krippendorff [40] were applied to ensure reliability, validity, and replicability in identifying patterns and themes. The data from the strategic documents was systematically coded and analysed using NVivo 12 software, which enabled the identification of critical themes such as digital integration, sustainability, community engagement, and resilience. The coding framework was developed based on research questions and existing literature, allowing for a detailed exploration of how these themes were emphasised and interconnected across the documents.

Third, to assess the relationship between the two strategies, a comparative analysis was conducted using NVivo's matrix coding query to identify thematic alignments and divergences between the two strategies. This method identified key areas of thematic alignment, such as shared commitments to environmental sustainability and collaboration

with stakeholders. It also highlighted differences in approaches to community engagement and technological priorities. Visual tools, such as Venn diagrams, were employed to illustrate the overlaps and gaps, providing a clearer understanding of strategic synergies and divergences. Additionally, narrative analysis was used to examine the language and framing of each document, uncovering implicit priorities and underlying strategic orientations [7,42].

The final step of Phase 1, Critical Analysis involved evaluating the strengths, weaknesses, and underlying assumptions of the strategic frameworks. This evaluation highlighted their strengths, such as a strong emphasis on resilience and inclusivity, as well as gaps and areas for improvement. For instance, while both documents prioritise stakeholder engagement, they offer limited consideration of marginalised communities and how technological advancements could address systemic inequalities. This detailed approach to analysing Sydney's strategic frameworks provided valuable insights into the city's urban policy and planning [43].

Together, these four forms of analysis provided a layered understanding of Sydney's strategic frameworks and uncovered both the explicit and the more nuanced aspects of urban policy and planning in Sydney.

3.3. Phase 2: Analysis of Smart City Projects

The second phase analyses a diverse range of smart city projects initiated or active between 2015 and 2023 in Sydney to understand how they contribute to the city's resilience and smart development goals. These projects were sourced from official city websites, government reports, academic publications, and specific project documentation, ensuring thorough coverage of Sydney's smart city initiatives.

This analysis utilised Typological Analysis to categorise and evaluate 36 smart city projects based on project type, scale, geographical focus, and ownership criteria. This systematic approach allowed for an assessment of how various project characteristics contribute to urban resilience. This approach was based on a framework developed by Bailey [44] to assess how well these projects align with the strategic goals revealed in Phase 1 and to reveal potential gaps that could hinder their effectiveness in fostering a resilient and sustainable urban environment.

Following this broad analysis, three key projects were examined in detail: the NSW Spatial Digital Twin, Land iQ, and SIMPaCT. These projects were selected for their significant impact on digital innovation, stakeholder collaboration, and urban resilience. Each project was evaluated using four different perspectives:

Technological Implementation: Evaluating the impact of AI, IoT, and big data and examining both the advantages and challenges.

Stakeholder Engagement: Analysing the interactions among public, private, and community stakeholders and their impact on project success.

Resilience and Sustainability Outcomes: Assessing each project's impact on Sydney's resilience and sustainability, focusing on real benefits and strategic progress.

Comparative Insights: Identifying each project's distinct and collective contributions to Sydney's overall smart city strategy, highlighting synergies and unique project advantages.

The textual data from both policy documents and project reports was processed methodically to ensure both consistency and reliability. NVivo 12 was utilised to assist with thematic clustering, keyword frequency analysis, and sentiment analysis. Additionally, visual representations such as matrices and diagrams were employed to synthesise the findings. These methods established a solid foundation for identifying patterns and drawing conclusions.

This study provides a comprehensive understanding of Sydney's smart city and resilience strategies through a combination of strategic and project-specific analyses. The findings reveal both successful integrations and areas needing improvement, enhancing our understanding of how urban systems can effectively align technology, governance, and resilience. This multidimensional approach not only advances academic discussions on

urban resilience but also offers actionable insights for policymakers and urban planners across various contexts.

4. Results

This section aims to provide a detailed analysis of the strategic documents that guide Sydney's smart city and resilience initiatives. It explores how technology and policy are integrated into these frameworks to foster resilient and innovative urban environments utilising various analytical methods. The results described below do not follow the five steps in the research approach (see above). Rather, it integrates the findings under three themes due to overlapping approaches. These themes are (i) core concepts in the strategic frameworks, (ii) typological patterns, and (iii) specific case studies.

4.1. Overview of Strategic Frameworks

This section synthesises the Descriptive and Content Analyses of Sydney's *Smart City Strategic Framework* and the *Resilient Sydney Strategy*. It highlights how these strategies integrate technology with policy and planning to promote a resilient, sustainable, and innovative urban environment.

4.1.1. Smart City Framework

The analysis of Sydney's *Smart City Strategic Framework* identified four crucial themes driving urban innovation and development in the city. These are:

Digital Integration and Infrastructure Development: The framework emphasises digital technologies such as IoT, AI, and big data analytics, improving service delivery and urban infrastructure and enabling efficient traffic management and predictive maintenance.

Sustainability and Resilience: Initiatives focus on environmental and social sustainability, creating green spaces, integrating sustainable practices into urban development, and bolstering a city's resilience against natural disasters and cyberattacks.

Inclusive Governance and Community Engagement: The framework promotes participatory governance and fosters public–private partnerships through digital platforms to promote community involvement and cross-sectoral innovation.

Equity, policy, and capacity building: To ensure fair access to technology, the strategy highlights comprehensive policy frameworks, data-driven decision-making, and skill development to maximise the benefits of smart technologies across all community sectors.

4.1.2. Resilience Strategy

The analysis of the *Resilient Sydney Strategy* identified key themes to strengthen the city's ability to withstand and adapt to urban risks and challenges.

Community and Infrastructure Resilience: Strengthening community networks and building resilient infrastructure are crucial to managing the risks and impacts of climate change and disasters. They ensure that social systems are strong and adaptable and that urban services and infrastructure (e.g., water and energy utilities and transport and communication networks) are hardened.

Economic and Environmental Sustainability: Resilience to disaster risk can be enhanced by promoting social and economic diversity and inclusion, supporting small and medium-sized enterprises (SMEs), and promoting environmental sustainability (e.g., with renewable energy projects). These activities support resilience by reducing vulnerability to sector-specific shocks, facilitating quicker recovery from disruptions, sustaining local economic development by enabling SMEs to innovate and reducing fossil fuel dependency. Moreover, they mitigate climate change impacts and improve the city's ability to absorb and recover from environmental shocks.

Comprehensive Emergency Preparedness and Health Initiatives: Developing early warning systems, disaster risk preparedness and prevention plans, comprehensive emergency responses and health services accessible to all and supported by digital literacy and continuous education to prepare residents with the skills to meet future challenges.

Data-Driven Leadership: Utilising advanced data analytics for effective governance and decision-making and aligning technology use with strategic urban management to foster a proactive, responsive urban system.

4.1.3. Integration and Synergies

Together, the themes in the two strategic documents provide a comprehensive approach to combining innovative smart city solutions with resilient strategies to create a sustainable, inclusive, and forward-looking urban environment. A comparison of the themes in each document reveals significant synergies, including Social, Economic and Environmental Sustainability, Community Engagement, Infrastructure Resilience, Digital Infrastructure, and Data-driven Decision-making.

Social, Economic and Environmental Sustainability: Initiatives such as the Sydney Green Grid and Greening Sydney Plan aim to enhance urban biodiversity, create recreational spaces, and improve public health. The Tech Startups Action Plan supports innovation and economic resilience, particularly among SMEs. Urban canopy mapping and greenhouse gas monitoring also leverage technology to manage environmental impacts, aligning with Sydney's sustainability goals and resilient urban environment.

Digital Infrastructure: Sydney is bolstering its resilience by strengthening its digital infrastructure. Initiatives, such as open data platforms and free Wi-Fi programs, enhance service delivery and ensure robust, flexible digital connectivity. This, in turn, facilitates innovation and enables community service continuity even during disruptions.

Community Engagement and Collaboration: Community participation in governance processes, such as participatory budgeting and disaster preparedness workshops, strengthens the city's resilience and fosters a collaborative, transparent, and engaged urban community.

Data-driven Decision-making: Real-time data analytics supports the management of Sydney's infrastructure and emergency responses, optimising city operations and enhancing proactive governance through predictive modelling and traffic flow optimisation.

Infrastructure Resilience: Investing in durable infrastructure, such as flood-resistant urban designs and upgraded bridges, aligns with ecological considerations and ensures readiness for future challenges. This provides a strong foundation for smart city technologies and resilience strategies.

The thematic connections between Sydney's Smart City Framework and Resilient Sydney Strategy are visually summarised in Figure 2. This diagram highlights the interrelations between the two strategic documents, emphasising their joint contribution to sustainable and resilient urban development.

4.1.4. Strategic Improvements and Recommendations

The above five interconnected themes promote the strategic integration of technology, community engagement, and resilient infrastructure planning to advance Sydney's smart and resilient city capabilities. However, Sydney's smart city and resilience strategies can be enhanced in six key areas to foster more adaptive and inclusive urban development. These areas are crucial for addressing observed gaps so Sydney can evolve into a smarter, more resilient city.

First, despite robust collaboration frameworks in the documents, stakeholder engagement has significant room for improvement, particularly for marginalised and underrepresented groups. Traditional methods, such as formal meetings and surveys, often fail to be accessible, potentially excluding certain groups due to language barriers, limited digital access, or a lack of awareness about engagement opportunities. To address this, creating inclusive and user-friendly feedback platforms and regular workshops focused on inclusivity and resilience can empower these groups to have a stronger voice in shaping urban strategies, ultimately promoting more equitable outcomes [45].

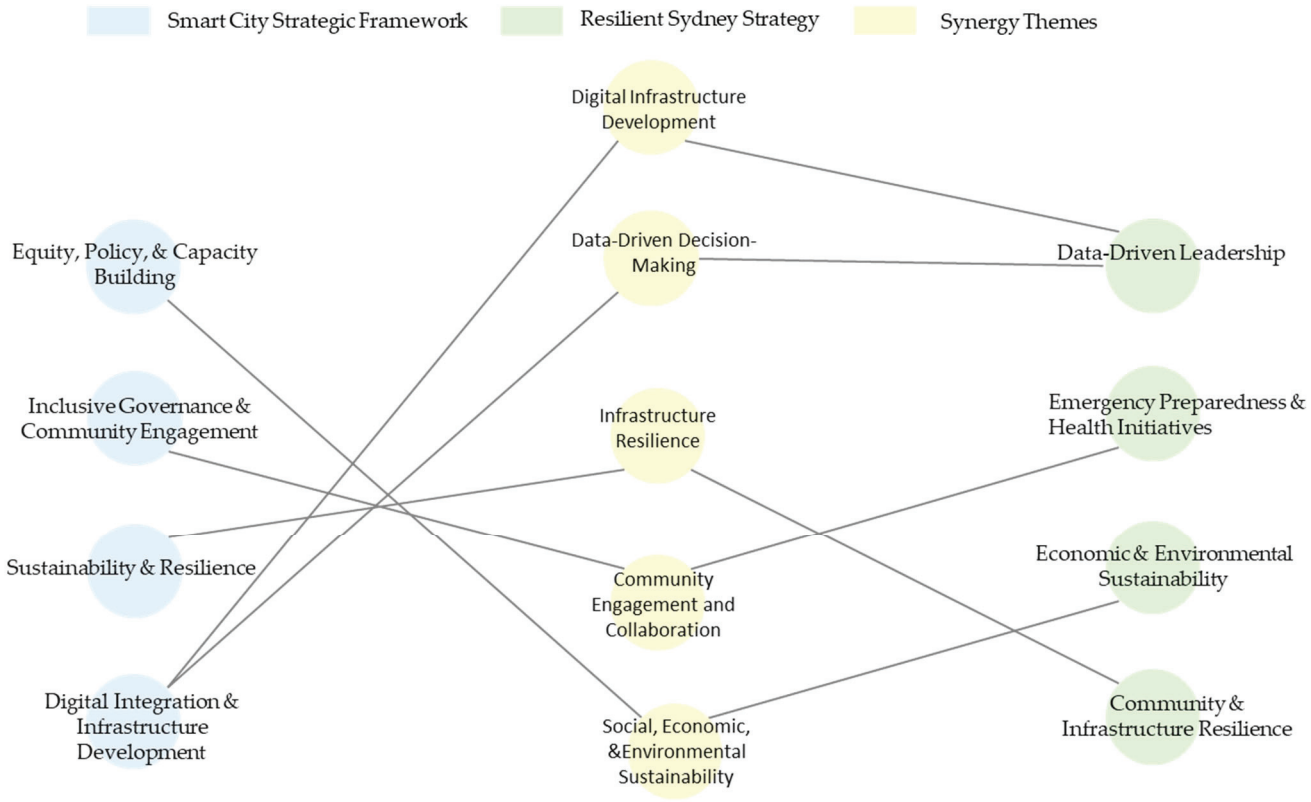


Figure 2. Integration of themes between Smart City Framework and Resilient Sydney Strategy.

Second, the equity considerations in the two strategies could be improved. For example, the inclusion of systematic equity assessments early in the project lifecycle can ensure that technological advancements and infrastructure developments are accessible to all socio-economic groups, thus avoiding any increase in urban disparities and, indeed, actively seeking to mitigate existing ones [46].

Third, the strategic documents could contain more specific implementation steps. Creating detailed plans with timelines, responsibilities, and measurable metrics can enhance strategic clarity and accountability, ensuring the initiatives' impact is both significant and transparent [47].

Fourth, the documents lack reference to the need to formalise replicating and scaling successful projects. Developing a team to capture best practices and assess scalability ensures that successful innovations are adapted to diverse urban settings and maximised for their benefit [48].

Related to this, a fifth improvement may be to strengthen frameworks and processes for monitoring and evaluation frameworks, especially to support agile responses to emerging challenges. For example, creating a digital platform for real-time reporting and analysis could significantly improve urban decision-making processes by allowing rapid adaptations to meet evolving needs [49].

Finally, Sydney's transformation into a resilient, forward-thinking urban environment will depend on adopting emerging technologies. Innovative solutions like AI and blockchain can enhance the efficiency and trust of public services, especially in election systems and property registration.

As a result of these strategic enhancements, Sydney can address its current urban challenges and create a solid foundation for anticipating and mitigating future challenges, thereby romoting sustainable and resilient urban development.

4.2. Typological Analysis of Sydney's Smart City Initiatives

This section reports on Phase 2 of the study, which sought to analyse the defining characteristics of Sydney's diverse smart city projects and their contribution to urban development and resilience. This Typological Analysis examines 36 projects using five key dimensions: project type, geographic focus, scale, ownership type, and resilience focus. The 36 projects were selected to represent Sydney's multifaceted approach to urban development and technology. This analysis is summarised in Appendix A.

4.2.1. Project Type

Sydney's smart city initiatives are divided into four distinct project types, each contributing to a dimension of the city's resilience and liveability. Figure 1 shows the proportion of each of the four project types.

1. Mobility projects constitute 41.6% of the initiatives and use solutions such as pedestrian sensors and 10 min neighbourhoods to enhance urban mobility. With the integration of IoT and AI for real-time traffic management, these strategies are key to reducing traffic congestion, promoting sustainable transportation, and improving city-wide accessibility (see Figure 3).

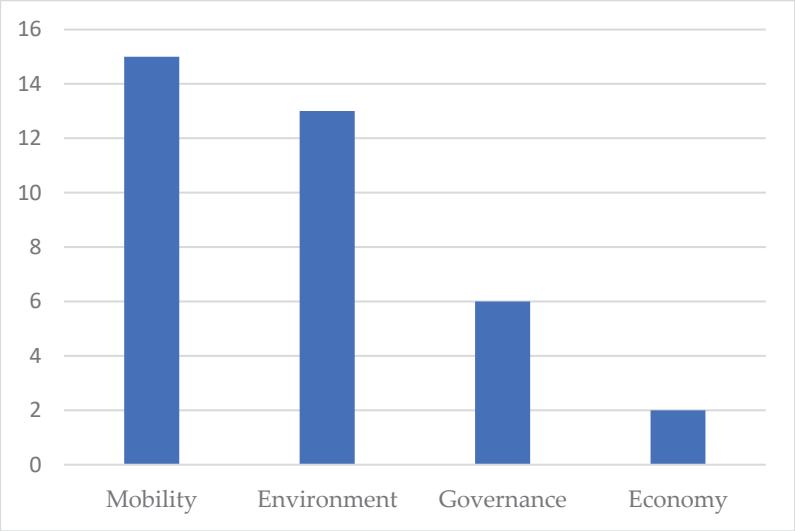


Figure 3. Distribution of smart city project Types in Sydney.

2. Environmental sustainability was the focus of 36% of the projects. These included urban canopy mapping and monitoring greenhouse gas emissions as key components of Sydney's climate resilience strategy. The initiatives leverage advanced sensor technology and data analytics to provide critical data for policy formulation and sustainable urban planning.
3. Governance projects constitute 16.6% of the portfolio and include projects aimed at improving civic engagement and operational efficiency by democratising data access, improving transparency, and facilitating participatory governance through Open Data Platforms.
4. Economic development makes up 5.5% of the initiatives but is crucial in stimulating economic growth and innovation. Efforts like the Tech Startups action plan and innovation districts foster entrepreneurship and capitalise on Sydney's technological and creative strengths.

4.2.2. Geographic Focus and Project Scale

In Sydney, smart city initiatives span different scales, addressing local and regional issues, as shown in Figure 4. About 60% of the smart city initiatives, such as the Open

Data Platform and city-wide Free Wi-Fi, operate at the city level, providing the essential digital infrastructure for real-time data collection and improved public service accessibility. Through these foundational services, seamless connectivity can be enabled, and various other initiatives can be supported.

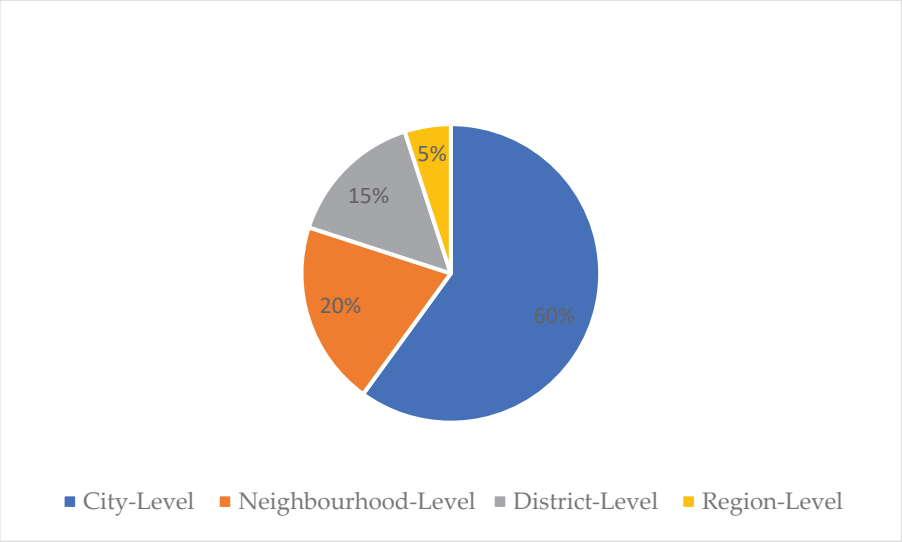


Figure 4. Proportional distribution of Sydney's smart city projects by scale.

Neighbourhood-scale projects account for 20% of the projects and focus on improving urban life quality, such as the 10 min neighbourhood initiative, which focuses on walkability and improving local access to services.

Projects at the district level, representing 15% of the initiatives, include Tech Central and Westmead Health & Innovation District, designed to promote technological innovation in health and economic growth within urban sub-regions.

Regional initiatives are being implemented in the remaining 5% of projects, with the NSW Spatial Digital Twin being one example. Using 4D visualisation, this project extends its impact beyond Sydney to enhance regional planning and management.

4.2.3. Project Ownership

Various ownership models support Sydney's smart city initiatives. The public sector is responsible for 45% of the initiatives, including the Open Data Platform and the NSW Spatial Digital Twin. Government-led projects contribute to improved urban management and governance.

Hybrid public–private projects represent 40% of the initiatives, combining public oversight with private-sector finance. Examples include the Tech Startups action plan and the Western Sydney 4D Digital Twin.

15% of the portfolio is devoted to private sector initiatives that offer specialised, innovative solutions tailored to specific urban challenges, such as intelligent parking systems and smart lighting solutions. A key feature of these projects is that they often incorporate cutting-edge technology and agile methodologies, such as AI-driven analytics and IoT-enabled sensors, to improve efficiency and adaptability in response to real-time data.

4.2.4. Resilience Abilities

Sydney's smart city initiatives focus on four key resilience abilities, often with several of these included in a project. The four are planning, adaptation, absorption, and recovery. These categories are based on well-established frameworks for urban resilience [10] and operationalised through thematic analysis of project objectives and outcomes.

Approximately 97% of the initiatives involve land use planning, which involves anticipating and preparing for future challenges and demands. A prime example is the NSW Spatial Digital Twin, which uses advanced modelling techniques.

The city's ability to adapt, evident in 72% of the initiatives, demonstrates its ability to adapt to environmental, economic, and technological changes. Smart Green Apartments, for example, enhance building efficiency to mitigate climate impact.

Around 47% of projects contain absorption capabilities, which enhance Sydney's resilience by minimising shock impacts immediately. An example of this resilience is Urban Heat Island Effect Mitigation.

Approximately 39% of projects incorporate recovery strategies to help them rebound quickly after a disruption. A data hub, Resilient Sydney, plays a crucial role in emergency response coordination, thus minimising long-term disruptions and restoring normalcy quickly (see Figure 5).

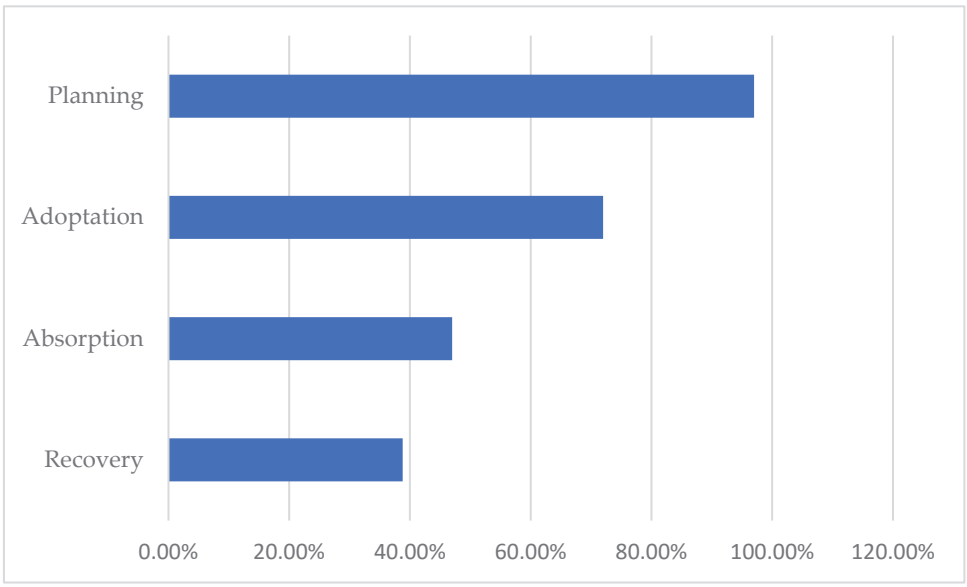


Figure 5. Distribution of Sydney's smart city projects based on approaches to resilience.

The typological analysis reveals the diversity and strategic integration of Sydney's smart city initiatives across various domains and scales. This analysis categorises projects into types, geographic focus, scales, ownership models, and resilience abilities, highlighting Sydney's commitment to using new technologies for urban enhancements and highlighting synergies among them.

The following section examines the direct contributions and operational dynamics of the NSW Spatial Digital Twin, Land iQ, and SIMPaCT within Sydney's urban context. Detailed examination of these technologies and strategies will reveal their effectiveness and impact on Sydney's vision for a smarter, more resilient city.

4.3. Project-Specific Analyses

This section evaluates Sydney's strategic progress in the smart city realm by analysing three pivotal projects: the NSW Spatial Digital Twin, Land iQ, and SIMPaCT. These projects were selected for their innovative integration of digital innovation, stakeholder collaboration, urban resilience, and the enhancement of sustainability and liveability. These demonstrate how Sydney uses advanced technologies such as AI, IoT, and big data analytics to address urban risks and challenges by improving its resilience, efficiency, and adaptability. Using technologies to such ends can improve the governance and sustainability of urban areas and ensure that the city is liveable and resilient under various stresses and hazards.

4.3.1. NSW Spatial Digital Twin: Urban Management and Resilience

The NSW Spatial Digital Twin was developed through a collaboration between government bodies and technology firms and utilises 3D and 4D modelling, IoT, AI, and machine learning to revolutionise how Sydney manages its urban infrastructure. This platform offers real-time, precise visualisations of the city, which allow planners to predict the effects of developments and simulate scenarios, such as traffic flows and environmental changes, before implementing them [50]. This aids in optimising strategic planning and enhancing emergency preparedness, which are crucial components of Sydney's Smart City and Resilient Sydney strategies.

The Spatial Digital Twin aligns with Sydney's smart city goals by enhancing data-driven decision-making and enabling digital integration. It aims to improve service delivery and foster more adaptive, efficient urban systems. Additionally, it can simulate the drivers and impacts of disruptions, thereby strengthening infrastructure resilience. This allows Sydney to proactively plan for and respond to disasters and infrastructure challenges, both of which are a central focus of the Resilient Sydney Strategy.

The platform prioritises sustainability by integrating environmentally conscious planning, including managing urban green spaces and reducing emissions. Furthermore, the public availability of geospatial data promotes transparency in civic engagement and a collaborative approach to urban development, both of which are essential for community resilience.

The Spatial Digital Twin project is a public–private partnership that supports Sydney's smart city innovation and resilience-building efforts. This project demonstrates how smart technologies can advance resilience, sustainability, and community engagement, directly contributing to Sydney's broader smart and resilient city objectives.

4.3.2. Land iQ: Innovating Urban Planning

Land iQ is an interactive database that integrates over 180 datasets and 40 land use typologies to support data-driven urban planning in New South Wales. Land iQ can model the environmental, social and economic impacts of development proposals, such as changes in water flow, biodiversity, ground cover, traffic patterns, access to public services, etc. The platform thus enables dynamic scenario analysis and strategic decision-making in support of sustainable urban development as well as provides planners with valuable insights that can strengthen urban resilience and contribute to the Resilient Sydney Strategy [51].

The platform also prioritises community engagement by making planning data publicly accessible. This encourages civic participation in shaping the city's future and promotes community resilience. Inclusive governance is a key objective of Sydney's smart city and resilience strategies.

Developed by Property and Development NSW and two private sector firms, WSP Australia and Giraffe, which are project management and software firms, respectively, Land iQ serves as a model of public–private partnerships in driving sustainable development and resilience through data integration and community collaboration. This makes it a key contributor to Sydney's Smart City and Resilient Sydney frameworks.

4.3.3. SIMPaCT: Enhancing Urban Environments

SIMPaCT (Smart Irrigation Management for Parks and Cool Towns) utilises AI, IoT, and environmental monitoring to optimise water usage in Sydney's parks by conserving resources and improving urban microclimates [52]. This system uses real-time data to facilitate decisions about reducing water waste and keeping parks cooler during hot months, directly supporting the sustainability and data-driven decision-making goals of Sydney's *Smart City Strategic Framework*. By mitigating the urban heat island effect and conserving water, SIMPaCT enhances Sydney's environmental resilience, thus also aligning with the *Resilient Sydney Strategy*. These efforts help the city adapt to climate-related challenges, which is a key focus of both strategies.

By keeping parks cooler, SIMPaCT aims to increase public use of these areas and encourage social engagement. This enhances community resilience by promoting social interaction and well-being in public spaces, aligning with the focus on strengthening community cohesion in the *Resilient Sydney Strategy*.

Supported by a public–private partnership between the NSW Government, Western Sydney University, and technology firms, SIMPaCT exemplifies smart environmental management and continuous innovation, which are central to both Sydney's Smart City and Resilience frameworks. It demonstrates how smart technologies can enhance urban resilience, sustainability, and community well-being, furthering the objectives of Sydney's Smart City and Resilient Sydney strategies.

4.3.4. Comparative Insights

Table 1 is a comparison of the characteristics and goals of NSW Spatial Digital Twin, Land iQ, and SIMPaCT projects. The table highlights each project's key features in terms of technological implementation, stakeholder engagement, resilience outcomes, and community impacts, highlighting its contribution to Sydney's smart city initiatives.

Table 1. Comparative Table of Project Outcomes.

Project Name	Technological Implementation	Stakeholder Engagement	Resilience Outcomes	Community Impacts
NSW Spatial Digital Twin	3D/4D visualisation of urban environments in real-time, IoT and AI integration for urban planning.	Collaboration between the public, private, and academic sectors improves data integration and decision-making.	Emergency management capacity enhanced, predictive maintenance supported, and infrastructure resilience enhanced.	Increasing civic engagement through geospatial data accessibility.
Land iQ	Comprehensive planning with over 180 datasets integrated Support for scenario planning in real-time.	Collaboration between the public and private sectors enhances strategic planning and implementation, integrating diverse datasets for comprehensive planning.	Achieved adaptive land use planning and contributed to sustainability and mitigation during environmental challenges data.	Increasing transparency and stakeholder involvement in urban development decisions.
SIMPaCT	A real-time environmental data-driven AI algorithm for irrigation optimisation.	Collaboration between communities and government promotes local ownership and tailor-made implementation.	Improved microclimates in urban parks due to optimised water usage.	Improved community well-being during heatwaves through enhanced public spaces.
Summary	NSW Spatial Digital Twin has the largest impact and is widely applied to urban management. SIMPaCT benefits environmental management through technology innovation, and Land iQ delivers strategic benefits in land use planning.	Extensive cross-sector engagement strategies exist for Spatial Digital Twins and Land iQs, resulting in scalability and urban impact. By contrast, SIMPaCT's localised approach effectively mobilises community resources and enhances public space management.	The projects highlight the importance of advanced digital tools for enhancing urban resilience. NSW Spatial Digital Twin excels in emergency management applications, Land iQ in strategic urban planning adaptability, and SIMPaCT in green space sustainability management, all of which boost urban efficiency and sustainability.	NSW Spatial Digital Twin increases public participation through data accessibility. Land iQ promotes transparency and public involvement in planning. SIMPaCT improves community well-being by enhancing urban parks, making them cooler and more inviting during heatwaves.

The findings of this study provide direct answers to the research questions by showing how Sydney's *Smart City Strategic Framework* and *Resilient Sydney Strategy* utilise advanced technologies like IoT, AI, and big data. In the thematic and typological analyses, significant synergies are revealed between these frameworks, particularly in promoting sustainability, improving infrastructure resilience, and enabling data-driven governance. There are, however, discrepancies in stakeholder engagement and implementation strategies, which pose challenges to realising these objectives. This detailed examination of three key projects—NSW Spatial Digital Twin, Land iQ, and SIMPaCT—demonstrates their transformative potential for fostering adaptive urban systems, but it also highlights gaps in inclusivity and scalability. It is clear that more robust mechanisms are needed for aligning strategic objectives with practical outcomes, especially when it comes to ensuring social equity and long-term sustainability. The following discussion critically evaluates these findings, situating them within the global smart city and resilience trends to explore their broader implications and actionable recommendations.

5. Discussion

This study has critically examined Sydney's *Smart City Strategic Framework* and the *Resilient Sydney Strategy*. It focuses on key projects such as the NSW Spatial Digital Twin, Land iQ, and SIMPaCT to understand how IoT, AI, and big data contribute to urban resilience. These technologies have been effective in fostering inclusive development, improving governance in the face of rapid technological changes, and ensuring compliance with data protection regulations, thereby aligning Sydney with global smart city trends [20,53].

5.1. Integration of Technology and Innovation

Sydney's strategic frameworks effectively leverage technology to tackle urban challenges, including environmental sustainability, mobility and community engagement. Projects such as the NSW Spatial Digital Twin offer predictive capabilities for urban planning, while SIMPaCT shows how real-time environmental data can improve water conservation and reduce urban heat island effects. These initiatives highlight the transformative potential of AI, IoT, and big data in developing adaptive urban systems that respond to evolving urban needs.

Recent advancements in smart city technologies highlight the importance of integrating geospatial intelligence and AI to enhance the efficiency and scalability of city management [8,14]. For instance, digital twin platforms are increasingly used to model urban dynamics, improve decision-making, and optimise resource use [54]. To stay at the forefront of innovation, Sydney must embrace emerging technologies such as blockchain for transparent governance and machine learning for climate risk modelling [30]. Expanding these capabilities can help Sydney transition to a proactive urban environment that anticipates and mitigates risks before they escalate.

5.2. Community Engagement and Stakeholder Collaboration

A key component of Sydney's strategies is community engagement; however, there are still gaps in the effective inclusion of marginalised and underrepresented groups. While platforms like Sydney Your Say promote public participation, recent research indicates a need to move from simple consultation to more collaborative co-creation models. Such approaches ensure that public input meaningfully shapes technological initiatives and urban policies [55,56]. For example, strategies such as gamified citizen engagement platforms or localised workshops could empower communities, particularly those facing barriers to digital access or socio-economic inclusion [14].

Global trends in digital equity emphasise the necessity of universal access to smart technologies. A targeted digital literacy strategy designed for Sydney's diverse population could assist in bridging this gap, promoting inclusivity and resilience throughout the community [45].

5.3. Managing Policy and Governance Challenges

The relationship between technology, policy, and governance is crucial to Sydney's smart city strategies. However, the rapid advancement of technology presents challenges in aligning policy goals with governance structures. For example, the NSW Spatial Digital Twin highlights the need for regulatory models that can adapt as technology evolves. To address these challenges, Sydney should focus on developing agile governance frameworks that allow for ongoing evaluation and improvement. Additionally, these frameworks should include dynamic data protection protocols to ensure user privacy is maintained in accordance with international standards such as the General Data Protection Regulation GDPR [8,55,57].

Global trends highlight the importance of collaborative governance, where the public, private, and non-profit sectors work together to develop policies that address common challenges. By embracing these principles, Sydney can improve its ability to balance innovation with accountability, ensuring the responsible and inclusive implementation of smart city technologies [12].

5.4. Committing to Environmental Sustainability and Urban Resilience

Sydney's dedication to environmental sustainability is clear through initiatives such as urban canopy mapping and real-time greenhouse gas monitoring. These efforts integrate nature-based solutions within smart city frameworks to address environmental challenges and enhance ecological resilience [34,58]. Implementing sustainable practices such as renewable energy systems, green infrastructure, and sustainable transportation solutions can reduce environmental footprints while enhancing urban resilience to climate risks.

A comprehensive sustainability model that addresses environmental, social, and economic aspects is critical for maximising impact. Integrating renewable energy systems like decentralised solar grids with urban planning initiatives can help reduce vulnerability to climate shocks while promoting economic inclusivity [32,59]. Moreover, additional research is necessary to assess how Sydney's smart city initiatives contribute to climate adaptation and social equity, ensuring that technological innovations mitigate systemic inequalities and provide shared benefits [31].

5.5. Limitations and Future Research

This study offers valuable insights, but it also has limitations. The reliance on case studies specific to certain projects restricts our ability to generalise the findings to other urban settings. Future research should involve comparative analyses across various global cities to extract broader lessons. Additionally, while qualitative data provides in-depth insights, it is essential to incorporate quantitative metrics as well. These could include measuring the reduction in response times attributed to Digital Twins or assessing the social impact of community engagement platforms for more thorough evaluations [22].

Future research should focus on the long-term impacts of smart technologies on urban resilience, examining how these technologies influence adaptive capacities over time. Additionally, studying how to scale successful innovations, like SIMPaCT, across other districts in Sydney would provide valuable insights into their replicability and broader applicability. It is also essential to address the digital divide to ensure that marginalised communities equally benefit from advancements in smart city technologies, thereby promoting inclusivity and equity in urban development.

By addressing these gaps, Sydney can enhance its status as a global model of resilient, inclusive, and sustainable urban development, establishing a benchmark for cities around the world to follow.

6. Conclusions

This study analyses Sydney's smart city and resilience strategies, focusing on the NSW Spatial Digital Twin, Land iQ, and SIMPaCT initiatives. These initiatives integrate technological, social, and environmental dimensions. The study examines how they lever-

age advanced technologies such as IoT, AI, and big data to enhance urban sustainability, promote community inclusion, and develop adaptive urban systems.

It is evident that Sydney's strategies have significant synergies. Using real-time data and AI-driven tools can enhance urban planning, promote sustainability, and encourage community engagement. With the Spatial Digital Twin, for example, real-time data integration can support predictive urban management, while SIMPaCT shows how AI and IoT can improve public well-being and improve climate resilience. In these examples, we can see how technology can play a transformative role in enabling adaptive and inclusive urban environments.

Despite the advancements in smart city initiatives, the study highlights several significant challenges. Current strategies often fail to effectively engage marginalised groups and ensure equitable access to the benefits of smart cities. Additionally, the lack of scalable frameworks to replicate successful initiatives, such as SIMPaCT, limits their broader application across different districts. Furthermore, mechanisms for evaluating long-term impacts and social outcomes are still underdeveloped, making it difficult to achieve Sydney's ambitious goals for inclusivity and resilience. Addressing these gaps presents opportunities for improvement and growth, positioning Sydney to become a global leader in smart city development.

To build on its successes, Sydney must improve stakeholder inclusivity by developing frameworks that actively engage underrepresented communities in the planning and implementation of smart city initiatives. Local workshops, gamified engagement platforms, and targeted digital literacy programs can effectively bridge participation gaps and promote social equity. Additionally, establishing formal processes to scale successful initiatives, such as SIMPaCT, across different urban areas will ensure that the benefits of innovation are accessible to larger populations while addressing local needs.

Agile governance models are crucial as technology advances more quickly than traditional regulatory structures can adapt. Sydney can address this challenge by implementing governance frameworks that promote continuous evaluation, iterative improvement, and the integration of emerging technologies such as blockchain and machine learning. Furthermore, it is essential to establish strong mechanisms for monitoring and evaluation to assess the performance and social impacts of smart city initiatives. This will help ensure alignment with strategic objectives and prepare for any unforeseen challenges.

Lastly, comprehensive sustainability must be emphasised. Incorporating nature-based solutions, renewable energy systems, and green infrastructure into Sydney's smart city framework will enhance ecological resilience, promote economic inclusivity, and reduce climate vulnerabilities. These initiatives should prioritise the intersection of environmental, social, and technological dimensions to maximise their impact.

Future research should expand the scope of this study by exploring the scalability and replicability of Sydney's smart city strategies in diverse urban contexts around the world. Longitudinal studies evaluating the long-term effects of these technologies on resilience, inclusivity, and sustainability will be essential. Additionally, it is important to address the digital divide and promote universal access to the benefits of smart cities in future initiatives.

By prioritising inclusivity, scalability, and adaptive governance, Sydney can strengthen its position as a global leader in smart city innovation and resilience. This study provides valuable insights into how the integration of resilience frameworks with advanced technologies can lead to more inclusive, adaptive, and sustainable urban solutions. It serves as a roadmap that cities worldwide can follow for guidance.

Author Contributions: Conceptualisation, S.V., J.F. and L.I.; research design, S.V. and L.I.; data collection, S.V.; analysis, S.V., J.F. and L.I.; writing the first draft, S.V.; review and editing, J.F. and L.I.; supervision, J.F. and L.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analysed in this study. Summarised typology analyses are provided in Appendix A.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. Comprehensive Typology of Smart City Projects in Sydney.

Project	Project Focus	Geographic Focus	Project Scale	Ownership Type	Resilience Abilities
Open Data Platform	Governance	City of Sydney	City	Public	Planning Absorption Recovery Adaptation
4 Innovation Districts 1. Tech Central 2. Westmead Health and Innovation District 3. Macquarie Park 4. Central Coast	Economy	1 and 3 have a local focus within specific districts or neighbourhoods of Sydney. 2 and 4 have a broader geographic scope extending beyond the city to encompass specific regions or precincts.	1. District 2. District 3. Neighbourhood 4. Regional	Combination of private and public ownership.	Planning Absorption Adaptation
Tech Startups Action Plan	Economy	City of Sydney	City	Hybrid	Planning Adaptation
Urban Canopy Mapping Technologies	Environment	City of Sydney	City	Public	Planning Absorption Adaptation
Air Quality Index (AQI) Monitoring System	Environment	The entire state of New South Wales	Various scales	NSW Department of Planning, Public	Planning Adaptation
Citizen Science	Environment	Sydney	Different levels	Public	Planning Adaptation
10-min Neighbourhoods	Mobility	City of Sydney	Neighbourhood	Public	Planning
30-min Cities	Mobility	Greater Sydney	Neighbourhood	Public	Planning
Smart Pedestrian Sensing Network	Mobility	City of Sydney	Street	Public	Planning
NSW Spatial Digital Twin	Governance	The entire state of New South Wales	Regional and state-level	Public	Planning Absorption Recovery Adaptation
Free Wi-Fi	Mobility	Sydney	City	Public	Planning Recovery Adaptation

Table A1. *Cont.*

Project	Project Focus	Geographic Focus	Project Scale	Ownership Type	Resilience Abilities
Communications and Connectivity Networks (e.g., 3G/4G/5G, broadband, Wi-Fi and low-power wide-area networks [LPWAN])	Mobility	Sydney	City	Private Wi-Fi: Public	Planning Absorption Recovery Adaptation
Sensor Technologies for Monitoring (e.g., GHG emissions, real-time traffic congestion, monitoring temperature, noise)	Environment	Sydney	City	Hybrid	Planning Absorption
Data Integration and Analytics Platforms	Governance	Sydney	City	Hybrid	Planning Recovery Adaptation
User Interfaces and Delivery Channels	Governance	Sydney	City	Hybrid	Planning Recovery Adaptation
Western Sydney's 4D Digital Twin	Environment	Western Sydney	The Entire Region of Western Sydney,	Hybrid	Planning Recovery Adoption
NSW Smart Sensing Network (NSSN)	Environment	NSW	NSW	Hybrid	Planning Absorption Adaptation
Smart lighting	Environment	District Within City	Varied From Building, Street, District	Public	Planning Absorption
Smart Parking	Mobility	Sydney	City	Hybrid	Planning Absorption
Community Engagement Platforms (e.g., Sydney Your Say)	Governance	Sydney	City	Public	Planning Adaptation Recovery
The Opal Pay Payment System	Mobility	Sydney	City	Public	Planning Adaptation
Sydney Coordinated Adaptive Traffic System (SCATS)	Mobility	Sydney	City	Public	Planning Absorption Adaptation
Intelligent Congestion Management Program (ICMP)	Mobility	Sydney	City	Public	Planning Recovery Adaptation
Public Transport Information and Priority System (PTIPS)	Mobility	Sydney	City	Public	Planning Adaptation
Real-time Public Information Display (PIDs)	Mobility	Sydney	City	Public	Planning Adaptation

Table A1. *Cont.*

Project	Project Focus	Geographic Focus	Project Scale	Ownership Type	Resilience Abilities
Open Data Hub and Developer Portal	Governance	Sydney	City	Public	Planning Absorption Recovery Adaptation
Contactless Transport Payments (CTP)	Mobility	Sydney	City	Hybrid	Planning Absorption Recovery
Smart Grid and Energy Management 1 (smart metering)	Environment	Sydney	Varied from building to City	Hybrid	Planning Absorption Adaptation
Smart Grid and Energy Management 2 (demand response programs)	Environment	Sydney	Varied from building to City	Hybrid	Planning Absorption Recovery Adaptation
Smart Grid and Energy Management 3 (integration of distributed energy resources)	Environment	Sydney	Varied from building to City	Hybrid	Planning Absorption Recovery Adaptation
Smart Water Management (monitor water quality)	Environment	Hybrid	Varied from building to City	Hybrid	Planning Absorption
Smart Water Management 1 (detect leaks and water wastage)	Environment	Sydney	Varied from building to City	Hybrid	Planning
Smart Water Management 2 (automated irrigation systems)	Environment	Sydney	Varied from building to City	Hybrid	Planning Absorption Adaptation
Digital signage and wayfinding	Mobility	Sydney	City	Hybrid	Adaptation
Autonomous Vehicles (e.g., Driverless shuttle bus trial)	Mobility	Sydney	Districts within the City.	Hybrid	Planning Adaptation Recovery
Automated driverless metro trains	Mobility	Sydney	City	Public	Planning Recovery Adaptation

References

1. Wood, T.; Arbabi, H.; Mayfield, M. Extreme Climate Change Impacts on Urban Infrastructure and Support Systems. In Proceedings of the EGU General Assembly Conference Abstracts, Vienna, Austria, 23–28 April 2023; p. EGU–15799. [CrossRef]
2. Vaidya, G.; Bindra, P.; Kshirsagar, M.; Tamane, S.C. Privacy and security technologies for smart city development. In *Security and Privacy Applications for Smart City Development*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 3–23. [CrossRef]
3. Yigitcanlar, T. Smart cities: An effective urban development and management model? *Aust. Plan.* **2015**, *52*, 27–34. [CrossRef]
4. Angelidou, M. Smart cities: A conjuncture of four forces. *Cities* **2015**, *47*, 95–106. [CrossRef]
5. Scholl, H.J.; Scholl, M.C. Smart governance: A roadmap for research and practice. In Proceedings of the IConference 2014 Proceedings, Berlin, Germany, 4–7 March 2014; Available online: https://www.ideals.illinois.edu/items/47419 (accessed on 10 October 2024).
6. SmartCitiesWorld. City Profile: Sydney. 2020. Available online: https://www.smartcitiesworld.net/city-profile/smart-cities-reports/smartcitiesworld-city-profile--sydney-5954 (accessed on 1 January 2023).

7. Dowling, R.; McGuirk, P.; Gillon, C. Strategic or piecemeal? Smart city initiatives in Sydney and Melbourne. *Urban Policy Res.* **2019**, *37*, 429–441. [CrossRef]
8. Kitchin, R. The real-time city? Big data and smart urbanism. *GeoJournal* **2014**, *79*, 1–14. [CrossRef]
9. Rockefeller Foundation. 100 Resilient Cities. Available online: https://www.rockefellerfoundation.org/100-resilient-cities (accessed on 1 January 2023).
10. Meerow, S.; Newell, J.P.; Stults, M. Defining urban resilience: A review. *Landsc. Urban Plan.* **2016**, *147*, 38–49. [CrossRef]
11. Martin, G.T. Deep sustainability: Really addressing our endangered human future. In *Struggles and Successes in the Pursuit of Sustainable Development*; Routledge: London, UK, 2020; pp. 34–44.
12. Papadopoulou, C.-A. Technology and SDGs in smart cities context. In *Smart Cities and the un SDGs*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 45–58.
13. Nam, T.; Pardo, T.A. Conceptualizing smart city with dimensions of technology, people, and institutions. In Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times, College Park, MD, USA, 12–15 June 2011; pp. 282–291.
14. Mortaheb, R.; Jankowski, P. Smart city re-imagined: City planning and GeoAI in the age of big data. *J. Urban Manag.* **2023**, *12*, 4–15. [CrossRef]
15. Adi Bhaskara, J.; Nurmandi, A. Role of artificial intelligence in the Smart City: A bibliometric review. In Proceedings of the International Conference on Human-Computer Interaction, Online, 26 June–1 July 2022; pp. 589–596.
16. Iyer, S.K.; Kumar, H. Integrating Big Data Technologies for Sustainable Urban Development: Analyzing the Impact of Data-Driven Decision-Making on Green Infrastructure and Smart City Initiatives. *Int. J. Soc. Anal.* **2024**, *9*, 1–12.
17. Sharifi, A.; Khavarian-Garmsir, A.R.; Kummitha, R.K.R. Contributions of smart city solutions and technologies to resilience against the COVID-19 pandemic: A literature review. *Sustainability* **2021**, *13*, 8018. [CrossRef]
18. Mora, L.; Deakin, M.; Reid, A. Strategic principles for smart city development: A multiple case study analysis of European best practices. *Technol. Forecast. Soc. Change* **2019**, *142*, 70–97. [CrossRef]
19. Olaniyi, O.O.; Okunleye, O.J.; Olabanji, S.O. Advancing data-driven decision-making in smart cities through big data analytics: A comprehensive review of existing literature. *Curr. J. Appl. Sci. Technol.* **2023**, *42*, 10–18. [CrossRef]
20. Bibri, S.E.; Krogstie, J. Smart sustainable cities of the future: An extensive interdisciplinary literature review. *Sustain. Cities Soc.* **2017**, *31*, 183–212. [CrossRef]
21. Khatibi, H.; Wilkinson, S.; Eriwata, G.; Sweya, L.N.; Baghersad, M.; Dianat, H.; Ghaedi, K.; Javanmardi, A. An integrated framework for assessment of smart city resilience. *Environ. Plann.* **2022**, *49*, 1556–1577. [CrossRef]
22. Fariniuk, T.M.D.; Hojda, A.; de Moraes Batista Simão, M. Searching a Resilient City: A Study About Theoretical-Conceptual Joints Between Smart City and Urban Resilience. In *Urban Resilience: Methodologies, Tools and Evaluation: Theory and Practice*; Castillo, O.F.G., Ed.; Springer: Berlin/Heidelberg, Germany, 2023; pp. 1–15.
23. Bibri, S.E. On the sustainability of smart and smarter cities in the era of big data: An interdisciplinary and transdisciplinary literature review. *J. Big Data* **2019**, *6*, 25. [CrossRef]
24. Sharifi, A. Resilient urban forms: A review of literature on streets and street networks. *Build. Environ.* **2019**, *147*, 171–187. [CrossRef]
25. Nations, U. Transforming our world: The 2030 Agenda for Sustainable Development. 2015. Available online: https://sdgs.un.org/2030agenda (accessed on 28 November 2024).
26. Al-Mujahed, L.K.A. Smart Cities and Sustainable Development Goals: A Systematic Review and Future Research Directions. *J. Sustain. Econ.* **2024**, *1*. Available online: https://scholarsin.com/schinsec/article/view/18 (accessed on 10 October 2024).
27. Höjer, M.; Wangel, J. Smart sustainable cities: Definition and challenges. In *Proceedings of the ICT Innovations for Sustainability*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 333–349.
28. Martin, C.J.; Evans, J.; Karvonen, A. Smart and sustainable? Five tensions in the visions and practices of the smart-sustainable city in Europe and North America. *Technol. Forecast. Soc. Change* **2018**, *133*, 269–278. [CrossRef]
29. Küfeoğlu, S. SDG-11: Sustainable cities and communities. In *Emerging Technologies: Value Creation for Sustainable Development*; Küfeoğlu, S., Ed.; Springer: Berlin/Heidelberg, Germany, 2022; pp. 385–408.
30. Ahvenniemi, H.; Huovila, A.; Pinto-Seppä, I.; Airaksinen, M. What are the differences between sustainable and smart cities? *Cities* **2017**, *60*, 234–245. [CrossRef]
31. Emrouzeh, M.P.; Fleet, G.; Moir, R. Smart cities initiatives to examine and explore urban social challenges. In *Toward Social Internet of Things (SIoT): Enabling Technologies, Architectures and Applications: Emerging Technologies for Connected and Smart Social Objects*; Hassanien, A.E., Ed.; Springer: Berlin/Heidelberg, Germany, 2020; pp. 77–98.
32. Gracias, J.S.; Parnell, G.S.; Specking, E.; Pohl, E.A.; Buchanan, R. Smart Cities—A Structured Literature Review. *Smart Cities* **2023**, *6*, 1719–1743. [CrossRef]
33. Hassebo, A.; Tealab, M. Global models of smart cities and potential IoT applications: A review. *IoT* **2023**, *4*, 366–411. [CrossRef]
34. Lee, J.; Babcock, J.; Pham, T.S.; Bui, T.H.; Kang, M. Smart city as a social transition towards inclusive development through technology: A tale of four smart cities. *Int. J. Urban Sci.* **2023**, *27*, 75–100. [CrossRef]
35. Echebarria, C.; Barrutia, J.M.; Aguado-Moralejo, I. The Smart City journey: A systematic review and future research agenda. *Innov. Eur. J. Soc. Sci. Res.* **2021**, *34*, 159–201. [CrossRef]
36. Shelton, T.; Zook, M.; Wiig, A. The 'actually existing smart city'. *Camb. J. Reg. Econ. Soc.* **2015**, *8*, 13–25. [CrossRef]

37. Sharifi, A.; Allam, Z.; Bibri, S.E.; Khavarian-Garmsir, A.R. Smart cities and sustainable development goals (SDGs): A systematic literature review of co-benefits and trade-offs. *Cities* **2024**, *146*, 104659. [CrossRef]
38. City of Sydney. Smart City Strategic Framework. 2020. Available online: https://www.cityofsydney.nsw.gov.au/strategies-action-plans/smart-city-strategic-framework (accessed on 8 December 2024).
39. City of Sydney. Resilient Sydney Strategy. 2018. Available online: https://www.cityofsydney.nsw.gov.au/governance-decision-making/resilient-sydney (accessed on 8 December 2024).
40. Krippendorff, K. *Content Analysis: An Introduction to Its Methodology*; Sage publications: Thousand Oaks, CA, USA, 2018.
41. Yin, R.K. *Case Study Research and Applications*; Sage: Thousand Oaks, CA, USA, 2018; Volume 6.
42. Bryman, A. *Social Research Methods*; Oxford University Press: Oxford, UK, 2016.
43. Fairclough, N. *Critical Discourse Analysis: The Critical Study of Language*; Routledge: London, UK, 2013.
44. Bailey, K.D. *Typologies and Taxonomies: An Introduction to Classification Techniques*; Sage: Thousand Oaks, CA, USA, 1994.
45. Goodman, N.; Zwick, A.; Spicer, Z.; Carlsen, N. Public engagement in smart city development: Lessons from communities in Canada's Smart City Challenge. *Can. Geogr./Le Géographe Can.* **2020**, *64*, 416–432. [CrossRef]
46. Van Heerden, Q.; Karsten, C.; Holloway, J.; Petzer, E.; Burger, P.; Mans, G. Accessibility, affordability, and equity in long-term spatial planning: Perspectives from a developing country. *Transp. Policy* **2022**, *120*, 104–119. [CrossRef]
47. McQuire, S. Urban digital infrastructure, smart cityism, and communication: Research challenges for urban e-planning. *Int. J. E-Plan. Res. (IJEPR)* **2021**, *10*, 1–18. [CrossRef]
48. Kuguoglu, B.K.; van der Voort, H.; Janssen, M. The giant leap for smart cities: Scaling up smart city artificial intelligence of things (AIOT) initiatives. *Sustainability* **2021**, *13*, 12295. [CrossRef]
49. Al-Badi, A.; Sharma, S.K.; Jain, V.; Khan, A.I. Investigating emerging technologies role in smart cities' solutions. In Proceedings of the Re-imagining Diffusion and Adoption of Information Technology and Systems: A Continuing Conversation: IFIP WG 8.6 International Conference on Transfer and Diffusion of IT, TDIT 2020, Tiruchirappalli, India, 18–19 December 2020; Part II. pp. 230–241. Available online: https://ifip.hal.science/IFIP-AICT-618/hal-03744828 (accessed on 1 June 2024).
50. Government, N. NSW Spatial Digital Twin. Available online: https://www.spatial.nsw.gov.au/digital_twin (accessed on 5 January 2024).
51. NSW, P.a.D. Land iQ. Available online: https://www.dpie.nsw.gov.au/housing-and-property/our-business/advisory-and-transactions/land-iq (accessed on 5 January 2024).
52. Government, N. Smart Irrigation Management for Parks and Cool Towns (SIMPaCT). 2023. Available online: https://www.nsw.gov.au/business-and-economy/smart-nsw-case-study-library/listings/smart-irrigation-management-for-parks-and-cool-towns-simpact (accessed on 5 January 2024).
53. Chourabi, H.; Nam, T.; Walker, S.; Gil-Garcia, J.R.; Mellouli, S.; Nahon, K.; Pardo, T.A.; Scholl, H.J. Understanding smart cities: An integrative framework. In Proceedings of the 2012 45th Hawaii International Conference on System Sciences, Maui, HI, USA, 4–7 January 2012; pp. 2289–2297. Available online: https://ieeexplore.ieee.org/document/6149291 (accessed on 6 April 2024).
54. Sivarethinamohan, R.; Reddy, R.S. Digital Twin for Smart City Resilience and Solutions. In *Digital Twin and Blockchain for Smart Cities*; Tyagi, A., Ed.; Wiley: Hoboken, NJ, USA, 2024; pp. 605–619.
55. Fagerberg, J.; Hutschenreiter, G. Coping with societal challenges: Lessons for innovation policy governance. *J. Ind. Compet. Trade* **2020**, *20*, 279–305. [CrossRef]
56. Tomàs, M. The smart city and urban governance: The urban transformation of Barcelona, 2011–2023. *Urban Res. Pract.* **2024**, *17*, 588–605. [CrossRef]
57. Breuer, J.; Pierson, J. The right to the city and data protection for developing citizen-centric digital cities. *Inf. Commun. Soc.* **2021**, *24*, 797–812. [CrossRef]
58. Petit, E.P. Smart City Technologies plus Nature-Based Solutions: Viable and valuable resources for urban resilience. In *Smart Cities Policies and Financing*; Vacca, J., Ed.; Elsevier: Amsterdam, The Netherlands, 2022; pp. 377–398.
59. Naguib, I.M.; Ragheb, S.A. Achieving Sustainability in Smart Cities & Its Impact on Citizen. *Int. J. Sustain. Dev. Plan.* **2022**, *17*. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

urban science

MDPI

Article

Reality vs. Expectations in the Implementation of Urban Agricultural Projects—A Polycentric Governance Analysis

Ouiam Fatiha Boukharta [1,2], Loïc Sauvée [2], Leticia Chico-Santamarta [1,3], Fabiana Pena-Fabri [2] and Luis Manuel Navas-Gracia [1,*]

[1] TADRUS Research Group, Department of Agricultural and Forestry Engineering, University of Valladolid, 34004 Palencia, Spain; ouiamfatiha.boukharta@uva.es (O.F.B.); lchico-santamarta@harper-adams.ac.uk (L.C.-S.)
[2] InTerACT Research Unit, Institut Polytechnique UniLaSalle, 76130 Rouen, France; loic.sauvee@unilasalle.fr (L.S.); fabiana.fabri@unilasalle.fr (F.P.-F.)
[3] Global Impact Department, Harper Adams University, Edgmond, Newport TF10 8NB, UK
* Correspondence: luismanuel.navas@uva.es

Abstract: Population growth and urbanization are threatening food security. Urban agriculture is therefore a solution for urban food production and distribution. This paper investigates a multi-level governance framework to evaluate how local authorities implement their prescriptions at different levels of decision-making and objectives for urban agricultural projects and their role in building urban resilience. A qualitative assessment based on interviews and fieldwork over two periods in 2023 and 2024 was conducted with stakeholders from different entities in the Normandy Region of France, including the metropolis, the city, and two projects' presidents. The findings revealed a positive alignment on polycentric governance between different entities in terms of socio-economic integration, climate improvement, and nutritional diversity, all of which were achieved by the cases evaluated. Additionally, local authorities are seeking to achieve urban food self-sufficiency in order to reduce the scale of food imports, thus highlighting a limitation and challenging aspect of this study, given that urban areas are compromised by population density, limited space capacity, and the impermanence of projects. This investigation clearly shows that using this combined systematic approach of interviews and fieldwork provides an in-depth understanding of authorities' needs and assesses the existence of polycentric governance compliance across multiple units.

Keywords: urban agriculture; food security; city resilience; multi-level analysis; polycentric governance; qualitative research; case studies; France

1. Introduction

The first and foremost human need after air and water is food, and currently, agricultural production is at risk due to population growth and the scarcity of water resources worldwide [1]. Indeed, food is fundamental to human well-being and development, and sustainability is achieved when people have access at all times to the food they need for a normal, healthy life [2]. Food security is thus determined by food stability, availability, and access, and is linked to livelihood security [3]. The main drivers of current food insecurity are global weather variability, urbanization, and population growth, along with various other factors responsible for changing food consumption patterns [4].

As a result, today's world is a combination of tradition, modernity, and agriculture [5], and where urban agriculture (UA) is currently able to occur wherever humans can cultivate grains, even in the smallest part of the soil [6,7], thereby providing interactions and adaptation to an urban ecosystem [8]. As a matter of fact, besides producing food from these urban spaces, green and sustainable cities enable cleaner air, water, and streets, empowering their environmental, social, and economic outcomes [9,10]. For those reasons, UA is a way to support sustainable development goals, where it includes environmental

protection, ensuring health and nutrition, decreasing poverty, along with societal economic development [11], and where the purpose of this article is to examine these aspects through real cases of UA, considering the polycentric governance (PG) dimension, which is crucial for the proper development of these organizations.

PG is an approach in which a variety of stakeholders come together to formulate regulations adapted to local contexts. With this approach, stakeholders can freely participate in different decision-making, policy formulation, and rule enforcement spheres according to their needs and concerns [12]. To do so, both metaphorical pillars and institutional design parameters are used to show how they can be mutually combined to achieve desired goals and manage spillovers. Within the context of UA, PG is seen as the interaction between multiple stakeholders (farmers, local authorities, non-profit organizations, etc.) within various institutional settings and rules aimed at ensuring that, under different conditions, everyone benefiting from a given commons pays their fair share and decides on that commons.

In other words, when thinking about polycentricity, it should be considered how autonomous actors can, nevertheless, interact formally and informally with each other through cooperation to ensure polycentric order or equilibrium within the system [13]. Such interconnection is highly pertinent in the UA context, as decision-making mechanisms are closely linked at different levels of the urban community, encompassing local projects, district councils, and city administrative structures [14]. This interconnection requires a polycentric approach to effectively coordinate and support the various urban activities and their benefits throughout the city and its inhabitants, in line with aspects of sustainable development, to which the present investigation is contributing.

This paper is structured according to the following chronology: Section 2 describes the various concepts used in this investigation and their application for a better understanding, while Section 3 presents the methodology employed, along with the main inclusion and exclusion criteria allowed to better trace our interviews with the different actors at the various institutional and regulatory levels. Section 4 presents the results of this analysis, highlighting PG and its effects on the involvement and realization of these urban practices, presenting the various similarities, diversities, and obstacles present, as well as evaluating two real-life cases for a more comprehensive assessment. Finally, Section 5 places these results in a clearer perspective, enabling conclusions to be drawn in Section 6, in which an overview of the current situation is presented, along with some recommendations that should be followed for better implementation of the practices.

2. Conceptual Framework and Application

2.1. Urban Agriculture (UA)

UA is defined as any type of agro-industrial activity located in or around the city, intended to provide products and ecosystem services to residents [15], including social, environmental, and economic impacts [6,16]. The benefits include physical and mental health improvement, alleviation of social and economic problems, and community resilience [15]. The primary aim of UA is to build healthier, more sustainable, and resilient communities, and not necessarily to produce large quantities of food [17]. Currently, UA represents a small but important percentage of the food distribution system in cities, since few urban agricultural projects (UAP) aim to replace traditional food distribution, in the expectation of leading to food self-sufficiency for individuals or cities [18].

It is therefore important to recognize that UA is not the only solution to address food insecurity and improve accessibility to food, but rather a transfer of responsibility for policy-makers to expect and institutionalize urban farms to serve as subsistence or primary food environment production sites managed by and for low-income communities operating without external support [19], making it imperative that these projects should be well led and managed by local decision-makers and authorities [20].

Expanding UA in densely populated environments requires complex coordination between different stakeholders, such as local authorities, developers, residents, and private actors [21]. Moreover, UAP encounter significant challenges in terms of governance,

particularly due to the polycentric structure of the decision-making process [22]. In this sense, it is also necessary to underline the importance of collaborative and participatory governance mechanisms, which, in the absence of effective coordination, struggle to ensure strategic coherence between the multiple stakeholders [23].

2.2. Polycentric Governance (PG)

PG can be defined as a self-organizing governance system composed of mutual adjustment and multiple governing actors, decision-making centers, and political issues, along with the relationships between them to better synthesize the concept [24,25]. Indeed, polycentric systems have been conceived as consisting of numerous centers of authority interacting internally and across scales for a shared governance purpose, where they facilitate equal representation of different governance actors, encourage policy innovation and diffusion, and support flexibility through the rapid reconfiguration of policy networks in order to achieve specific goals [12,24], which is clearly explained through Figure 1. PG is therefore a vehicle for specialization, distribution of work between the central, regional, and local levels; subsidiarity; and adaptation of interventions to local–regional circumstances and community preferences, improving effectiveness in the context and scale of the specific challenge [26], which is the focus of this research in order to evaluate this departmental distribution of work between the various local and regional entities.

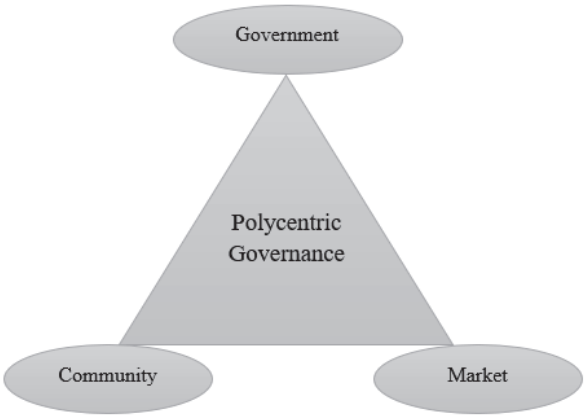


Figure 1. PG on the central Government-Market-Community axis.

PG can be decomposed into different elements, particularly governance actors, decision-making venues, policy issues, etc. [27]. Governance actors include individuals who influence decision-making processes and who have an interest in one or more policy issues or who are affected by decisions taken, with the aim of addressing coordination and conflict for multiple purposes [28] (Figure 1). Actors can include government agencies and civil servants, user groups, scientists and consultants, individuals from the general public, etc. In relation to the decision-making venue, this is a space of collective action with the power to design and adopt public policies and implement them, and where actors engage with each other, which enables actions to be initiated to make a joint decision [29]. In terms of the political issue, these are the topical policy areas or issues they address, and where issues can be both general and specific and vary in terms of importance and complexity. Considering these types of relationships, polycentric systems present multiple links through various actors who affect other places through their political decisions in the form of institutional rules and political decisions [26].

Regarding the PG model, according to Frimpong Boamah (2024) and the purpose of this paper, it enables a more integrative and adaptive management of urban resources [12], thus empowering diverse stakeholders to participate in the policy-making process that is crucial for UA and the socio-political dynamics that influence social patterns of urban food distribution and access. These interconnections within polycentric systems are therefore essential to fostering resilience and equity in sustainable food systems.

2.3. Scope of This Research Article

The aim of this study is to critically evaluate the alignment between metropolitan and urban policy frameworks and the practical implementation of UA initiatives. Indeed, it investigates the implementation and involvement of various UAP in France, more specifically in the metropolis of Rouen and its central city, applying a PG framework and an experimental field approach and focusing on their contribution to building urban resilience. It also considers the way in which these initiatives align with sustainable development goals and the transition to green cities, along with the impact of governance structures on their integration, coordination, and effectiveness within the framework of urban policies. These elements will be reinforced by the presentation of concrete UAP and their evolution over time between 2023 and 2024.

Such an analysis seeks to identify how multiple decision-making centers, such as local authorities, project leaders, and local community groups, can operate independently while simultaneously interacting together at different scales (local, regional, and global). These centers of authority work together to achieve shared governance, i.e., the collaborative management of decision-making and resources through cooperative action and mutual accountability, rather than relying on a single governing body.

3. Materials and Methods

The methodology followed in this paper consists of conducting semi-structured interviews with different actors, along with fieldwork analysis. Indeed, applying this approach makes it possible to control and monitor changes over time and to gain an overall understanding of the process [30].

3.1. Methods and Data Collection

Data were collected through semi-structured interviews, a qualitative method widely used to explore participants' perceptions and experiences at a deeper level, using an interview guide prepared by our research team, providing a clear, structured framework while allowing the order of questions to be adjusted so that the interviewer feels comfortable and flexible as the interview progresses [31], all selected using a purposive sampling approach [32,33]. In fact, this methodology enables an in-depth analysis to be provided at different scales and from several angles, including detailed qualitative information on interviewees' perspectives [34]. Moreover, systematic thematic analysis was used to pinpoint and interpret significant patterns in qualitative field interview data [35], enabling new and emerging concepts to be identified and developed, with rigorous double coding and in-depth analysis.

The interviews were conducted with 28 stakeholders, where $n = 15$ from the metropolis, $n = 10$ from the city, and $n = 3$ for project leaders. A multimodal approach was therefore used, consisting of face-to-face interviews and, in the case of non-compliance, telephone calls and/or online meetings, in order to adapt to participants' schedules and availability [36]. Such an approach ensured broad coverage of perspectives while respecting logistical constraints and stakeholder preferences [37]. Each interview was recorded with the participants' informed consent, in accordance with the principles of research ethics [38]. All interviews were transcribed in their entirety to guarantee the most accurate data analysis and processing possible. Fieldwork experimentation has also been performed in order to evaluate in a concrete way the UAP and to examine the evolution over time [39].

The experiment consisted of two phases, covering periods from 2023 to 2024, in which each phase tested the progress made over time and/or the performance/success of the projects, providing a more comprehensive perspective on project dynamics, encompassing more than concrete progress, challenges and strengths, and even the sustainability of initiatives [31].

3.2. Actors Involved

The aim of in-depth interviews is to include a variety of perspectives and expertise, thus fostering a comprehensive assessment of institutional aspirations and actual expe-

riences in the field [40]. This analysis covered several assessment dimensions, including the metropolis, the city, and real-life UA practices, all selected using a purposive sampling approach and a PG analysis to ensure relevance and diversity of perspectives, resulting in a multi-level governance (MLG) assessment [32,33], as illustrated in Figure 2.

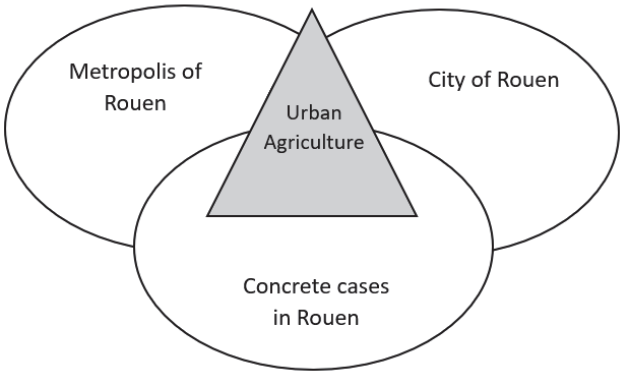


Figure 2. MLG analysis of UAP.

The MLG strategy provides generated, rich, narrative qualitative data, which are necessary for nuanced insight into systemic relations and social interaction, and where the outcomes are a very useful empirical basis for strategic recommendations and reinforcement of future urban planning [41]. Moreover, this methodology of MLG combined with a PG analysis is important in drawing attention to disparities and similarities between public policy and existing operational practices, which can enhance the ability to chart an effective pathway to plan urban transformation for sustainability [42,43].

Interviewees and Their Positions

Several actors occupying different positions were involved in this investigation in order to guarantee a PG approach and an MLG analysis of the context studied, as shown in Figure 3:

Actors from the three entities interviewed in this investigation		
Metropolis of Rouen	**City of Rouen**	**Real-life cases of Urban Agricultural cases**
Director of environmental transition: biodiversity, agriculture and green cities; Director of the Ecological Transition Department: Sustainable gardening and urban agriculture projects. Unit managers for sustainable gardening and urban agriculture: Promoting sustainable urban agriculture practices	Urban renewal project managers: Leads an urban renewal program to transform neighbourhoods and public spaces. Leaders of the plant sciences and ecology department: Activities and events around gardening. Biodiversity, ecology, resilience of green spaces, etc.	"Le champ des Possibles": - President of the garden: management of the garden to feed people, and to have awareness-raising around food. - General Delegate and Responsible for all operational aspects: Implementing the ambitions of the Administrative Council at association level. "Le Jardin de l'Astéroïde" - President of the garden: Management of the garden, contact with authorities and solving issues when needed

Figure 3. Key actors, their roles, primary functions, and presidents of urban practices in the metropolis and City of Rouen.

A total of 28 semi-structured interviews were conducted with local authority representatives, UAP managers, community leaders, and experts in urban planning and sustainable development (Figure 3). These interviews provide a better assessment of the coverage of the three entities towards the same objectives or whether there is divergence or non-complementarity. This analysis consist of evaluating the MLG perspective approach together with the PG framework in order to make it possible to understand the current situation of UAP through the expectations and real-life cases of UAP while anticipating and monitoring future evolution and changes.

3.3. Study Area

This investigation has been carried out in the Region of Normandy (Rouen metropolis and Rouen city), located in northern part of France.

3.3.1. Normandy Region

The French region of Normandy is widely known for its potential for agricultural production [44], partly unexplored in other studies, and for the ongoing urban support of local authorities from the different PG levels, including the metropolis, the city, associations, etc. The following Table 1 shows the main economic variables of this region, including population, agricultural area, gross domestic product (GDP), agricultural value added, etc.

Table 1. Key economic variables of the Normandy Region of France.

Economic Variable	Value	Reference
Population	3.3 million inhabitants	[44]
Area	29,906 Km^2	[45]
Agricultural area	22,000 Km^2	[46]
GDP	EUR 95 billion in 2018	[44]

Table 1 clearly shows that Normandy's agriculture covers 70% of its territory, making it a major component of the region's economic system. It is also, through its practices and diversity, an activity that shapes the Normandy landscape [46]. The local economy in Normandy is very strong, particularly in the agricultural and agri-food sectors, but still relatively modest compared with the rest of France, especially the Ile-de-France region [44]. However, to better assess UA practices in this area, it is necessary to apply a PG analysis across multiple actors and at different scales, assessing what is requested by local authorities and what is actually performed and applied in real UAP cases in the city.

3.3.2. Metropolis of Rouen (MR)

The MR Normandy is an intercommunal group of 71 communes that includes the CR [47] and covers most public transport (especially in Rouen) as well as managing waste, etc. The metropolis is a pivot point in the sustainable development of the region, where optimization is oriented towards quality of life [48], promoting innovative ideas like urban farms and community gardens that increase biodiversity along with recreational and educational opportunities for citizens [23].

3.3.3. City of Rouen (CR)

The CR is situated on the banks of the Seine River in France and is the historic capital of the Normandy Region [49]. The city is famous for its extensive cultural and historical heritage, featuring a number of emblematic monuments. Regarding UA, the CR is engaged in promoting initiatives to integrate green spaces and UAP, where community and allotment gardens are established to promote local and sustainable food production while strengthening social links and promoting environmental awareness [23].

3.4. Urban Agricultural Cases Evaluated

«Le champ des possibles» (CP)

A non-profit association located in two sites, one within the CR, the other in Sotteville city in the MR, and which is engaged in the development of UA and the ecological rehabilitation of derelict spaces in urban areas. Its main objective is to transform urban wastelands into areas dedicated to biodiversity and local agriculture and to ensure access to healthy food [23]. Moreover, awareness-raising, animation, and training activities around food, fostering a collective and participatory approach on the part of local residents, make this association a key reference for all other UAPs in the region. Such actions help to reconnect inhabitants with nature and encourage ecological practices in an urban framework [50]. To maximize the impact of its actions, the association collaborates with public institutions, schools, etc., and is supported by CR and the MR with the aim of enriching urban spaces with food production while strengthening social cohesion [51].

«Le jardin de l'Astéroïde» (JA)

Allotments which operates on a non-profit associative model and whose main objective is to promote agroecological practices in an urban environment while enabling residents to grow their own fruit and vegetables and participate in garden activities [26]. Prior to its creation, this site was used as an unauthorized parking lot. Its transformation into an allotment not only improved the value of the area but also contributed to environmental protection and the enhancement of biodiversity [52]. The garden is made up of small plots of 20 square meters and two large plots of 40 m^2, where each member has their own plot and is obliged to maintain and cultivate it properly while not being allowed to use chemicals or make any kind of sale [26]. The project is supported by the CR and works with other local associations to strengthen UA initiatives. This partnership makes it possible to create quality green spaces that meet the challenges of urban reappropriation while integrating an educational and green community dimension [53].

4. Results

Interviews and fieldwork enabled us to gather all the information needed to answer our problem, which is to study the interest of an MLG framework to assess the actual implementation of local authorities' mandates at different levels of decision-making devices, along with an assessment of the objectives of UA schemes and their role in strengthening urban resilience.

4.1. Stakeholders' Objectives

The interviews revealed a number of strategic objectives that encourage the development and promotion of urban areas. Figure 4 illustrates the main objectives that the metropolis, the city, and the two UA cases aim to achieve while ensuring compliance with the PG approach, emphasizing empowerment and decentralized decision-making.

From Figure 4 above, the MR asserted that the creation of these spaces is aligned with the French law on zero net artificialization. The effort hopes to reduce the future environmental impact and foster the re-naturalization of regions affected by urbanization at a rate of 66%. Furthermore, it is worth mentioning that this approach represents a concern in today's progress of the MR, signaling commitment to locally ecological sustainability, raising urban air quality and life conditions, and fostering food self-sufficiency/autonomy, which has been mentioned by all metropolitan respondents.

Regarding the CR, employability was highly mentioned by the stakeholders interviewed, with a rate of 66%, as it enables individuals to be inserted and/or reintegrated into the job market, thus emphasizing the human aspect and the social link. Furthermore, guaranteeing the social link remains a central objective that was mentioned by all our interviewees (rate of 100%), underlining the desire to integrate the notions of neighborhood and intergenerational interactions (Figure 4). Interveners from the CR also emphasize at 100% that one of their main objectives is to improve certain plots of previously unused fallow

land by giving them a cultivated character, thus promoting environmental sustainability and food self-sufficiency.

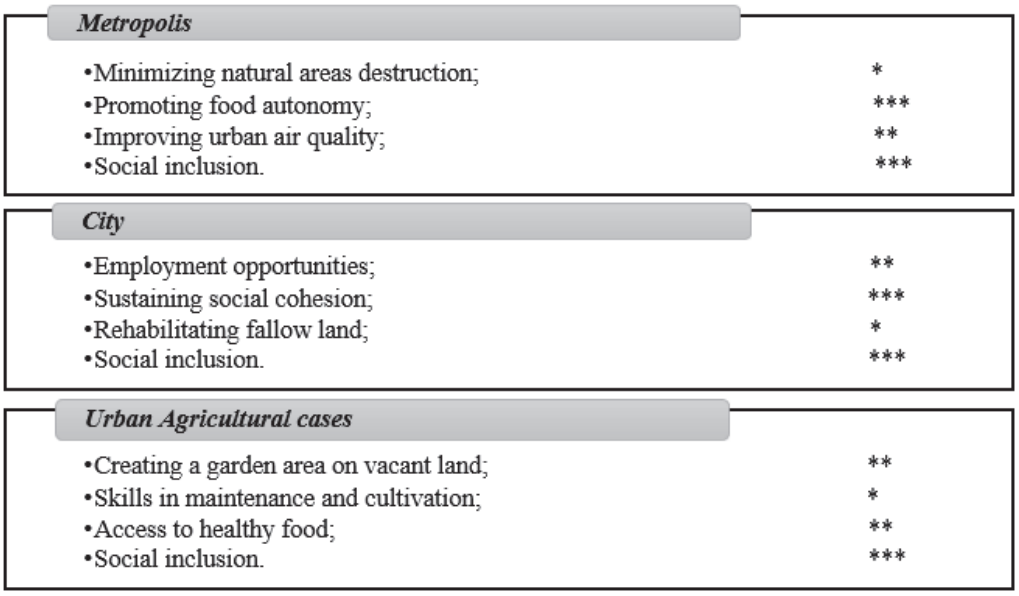


Figure 4. Main objectives that each of the metropolis, city, and UA cases in Rouen are aiming to achieve. The symbol * refers to the citation's intensity, where: * least cited; **: moderately cited, and ***: highly cited.

Last but not least, the presidents from the two real-life cases of UA evaluated stressed at a rate of 100% the importance of converting vacant lots into shared gardens to prevent them being used as parking lots or for other purposes by residents. This initiative aims to optimize the use of urban spaces while promoting the learning of the skills needed to maintain and cultivate the plots, thus guaranteeing participants' food self-sufficiency. In addition, they have expressed a desire to maximize the greening of the city while integrating a greater proportion of vegetation into the urban landscape. More broadly, this is a part of an integrated, beneficial approach that provides soil conservation and human wellness while building social bonds between neighbors to increase the connection and positive environment within the cities (Figure 4). However, only 33% of participants mentioned the need to have maintenance and cultivation skills developed beforehand, as for them, these skills are acquired over time.

4.2. Effect of UA on Residents

Interviews with different actors were performed to examine the effect of UA on local inhabitants. Figure 5 summarizes the main aspects cited by our interviewees, highlighting the effects of these urban spaces on residents.

Figure 5 highlights the impact of the insertion of UA practices on local residents and shows that there is considerable diversification between the entities examined, revealing both similarities and differences, all related to the three aspects of sustainability, categorized into social, economic, and environmental. Firstly, the social aspect is highlighted by all three entities with a rate of 100%, emphasizing its central role in facilitating interactions and exchanges between participants. In addition, this aspect contributes to improving the health of residents, fostering their social integration, personal development, and a reconnection with nature, which today tends to fade away in urban environments.

The second aspect is the food dimension, mentioned with a rate of 80%, where participants emphasize that gardening brings tangible satisfaction through the production and consumption of healthy, diversified food. Not only does this practice reinforce reconnection

with daily eating and food self-sufficiency, which also acts as an economic lever, helping to reduce expenses while maintaining a high-quality diet.

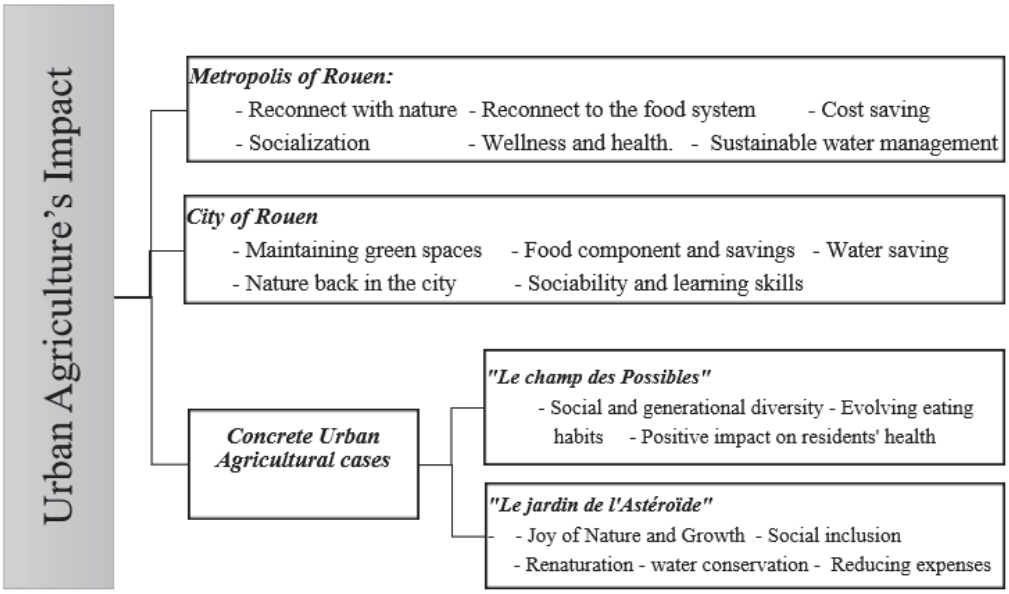


Figure 5. Impact of UA on local residents, according to the metropolis, city, and UA cases in Rouen.

Furthermore, from Figure 5, the third most frequently mentioned item relates to the preservation of biodiversity and water resources by promoting renaturation and protection of the environment, and thus, with a rate of 60%. Indeed, with the expansion of green spaces in urban areas, it is ensured that protected natural habitats are largely designed because of the additional space, and these areas act, therefore, as formidable water filters, increasing infiltration and enabling the replenishment of water tables.

4.3. Baseline Criteria (Indicators)

The creation or development of an urban area is subject to a number of requirements and indicators, prompting local authorities to choose one project over another. The local authorities evaluated share a number of common criteria, including soil analysis prior to implementation, the pursuit of food self-sufficiency, and the efficient use and preservation of water resources. In addition, other indicators have been mentioned and guaranteed by the UAP, as shown in Figure 6 below and explained below.

With regard to the common indicators mentioned by the three entities, our interviews reveal that local authorities (MR and CR) require soil analyses to be carried out beforehand, in order to guarantee two aspects. Firstly, to ensure that the land is suitable for the production of good quality fruit and vegetable crops, and secondly, to identify any missing nutrients so that they can be adequately compensated in the right way. The CP president explained that, to ensure the smooth running of the project, 40 cm of soil had to be removed and new soil had to be added, which is a measure imposed even though it had not originally been requested.

The second aspect concerns food autonomy (Figure 6), in which a key condition is that members have access to healthy, fair, and sufficient food. This aspect is indeed fundamental and must be achieved, whether partially or completely, as emphasized by the presidents of the UAP evaluated. The third common element involves sustainable use, protection, and management of water resources. In fact, the protection of catchment areas for drinking water is a crucial requirement long regulated and compelling local governments to make drastic arrangements to keep water resources healthy. These regulations are essential to meet the growing demands of urbanization and environmental constraints.

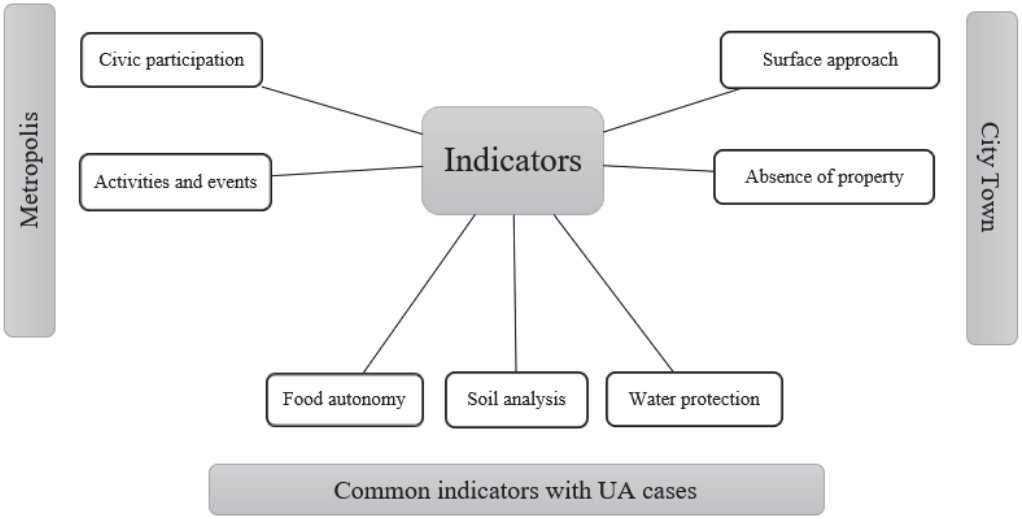


Figure 6. Criteria by entity that were mentioned by participants to ensure UAP.

Besides the three core criteria covered by the three distinct entities being evaluated, additional indicators were identified as priorities, highlighting the MLG perspective, particularly for MR and CR (Figure 6). The metropolis emphasizes the need for citizen engagement in the learning of agricultural practices and agronomic processes but also participates in various activities related to green community life that contribute to strengthening social ties. Some of the key metrics also include the number of events organized annually, annual visitors, and frequency of public access to these initiatives. On the other hand, the CR is concentrating on compliance with certain regulatory constraints, in particular maintaining a minimum surface area of 10 m^2 for UA spaces and ensuring that no real estate projects are planned on the land, in order to guarantee its long-term use.

4.4. Institutional and Collaborative Partnerships

According to our investigation, relationships between local authorities (city and metropolis) and UA project leaders, as revealed through interviews, are structured around regular exchanges and well-defined collaborations:

MR: At the outset, coordination takes place mainly with the project leaders, while the involvement of local residents takes place on the ground via public consultations. In addition, a service provider accompanies each project individually, with annual reviews to assess progress, as mentioned by a metropolitan authority stakeholder, "At the outset, our contacts are the promoters, where there is often preparation for the application. . . Then, there's the part where we set up individual support . . . And after that, we're back in touch with them for a review at the end of the first year".

CR: Regular feedback and consultations are organized to define and adapt projects according to needs. Partnerships with associations, such as support in signing the urban gardening charter, are formalized. A Rouen city official said, "Essentially there are three steps: first phase of consultation, second phase, of formalization and third phase we let the ground to the association.

In both cases, MR and CR representatives meet regularly with UAP presidents and members to monitor the impact of their project on the local population, as well as to follow up on signed agreements, thus promoting the accountability of those managing the initiative. By sharing responsibilities and providing a flexible framework, the unified action program can be successful and sustainable over the long term.

This institutional and collaborative management mode emphasizes the respected PG process, where various decisions are taken at all levels, thus promoting local autonomy and flexibility in management. However, it is necessary to integrate and respect the global rules imposed by the various local and regional authorities through an MLG perspective.

4.5. The UAP Evolution over Time

UAP were the subject of field visits in both 2023 and 2024, following the completion of the interviews carried out to follow up on their progress and evolution. At JA and CP, a mix of causes were responsible for noteworthy differences at both plots. In the case of JA, the CR financed the installation of fencing and supported the maintenance of various infrastructures, such as the management of wetlands and the repair of deteriorated parts of the garden. Lighting has also been improved with the installation of light poles. The most important transformation concerns the change in ownership of the project's presidency between now and 2023–2024, where the president (the founder of the project, itself) has ceded his position in favor of a loyal member who has been fully committed to the project from the outset.

A part of the park was also allocated for an agroforestry pilot project, named CP, to demonstrate that growing trees within a farm can reduce water consumption. Between the two visit periods, an information panel was installed at the entrance to raise public awareness of the project's objectives. Work is currently underway to extend the infrastructure and accommodate a wider public, while it is planned to double the number of activities, particularly workshops and events. The extension of the growing area is also part of a new operation in support of isolated pregnant women, funded by the MR. Besides the salaried employees already in the organization, the project's governing bodies have added a full-time coordinator, two full-time animators in 2023, and a master-trained gardener. Furthermore, new agreements have been signed to ensure the expansion and sustainability of these initiatives.

4.6. UAP's Alignment with the Authorities' Priorities

One of the main questions of this research study is to determine the relevance of UAP in the CR and its metropolis in meeting the needs and expectations of local authorities and to evaluate the effectiveness of the links and the interaction between the three entities. The interviewees' contributions provided answers to this question, revealing a wealth of information on each of these points, as shown in Figure 7.

According to the stakeholder responses presented in Figure 7, there is effective complementarity and coordination between the positive and negative responses in terms of meeting the needs identified. The MR emphasizes the importance of public health, as well as the organization of events and awareness-raising days to inform the public of the benefits of UA. It affirms that UAP fully meet these objectives and that the metropolis is generally satisfied with the results obtained while motivating local entrepreneurs to become actively involved, believing that there is always the opportunity to learn and do better. Urban project managers agree: "We are finally starting to address the underlying factors of a healthy society and most importantly food". Together with the CR, the local authorities confirm that the social and economic aspects are also satisfied, in that participants of all generations and nationalities are well integrated into the projects. Moreover, the local production of fruit and vegetables enables residents to make some savings, even if these remain modest.

However, as far as the unmet expectations of local authorities are concerned, it is noted that the capacity of UAP to feed the whole city remains limited. As the president of CP pointed out, "UA won't be able to feed the city". Nevertheless, it is already contributing to this objective, which is seen as a positive result. However, institutional stakeholders strongly suggest that UAP should aim to achieve exceptional initiatives, such as the introduction of innovative crops, in order to attract greater public interest and thus reinforce their impact (Figure 7).

These interactions occur in the context of PG, where the local government at different levels (metropolis and city council) and civil actors (project developers) coordinate to manage UAP. This governance model strongly encourages shared, decentralized management, in which each stakeholder can respond to specific local requirements while maintaining coordination of all practices. This form of multi-centric cooperation helps to increase the

flexibility and adaptability of urban systems to address local challenges such as public health and social integration, enhancing the resilience of urban systems.

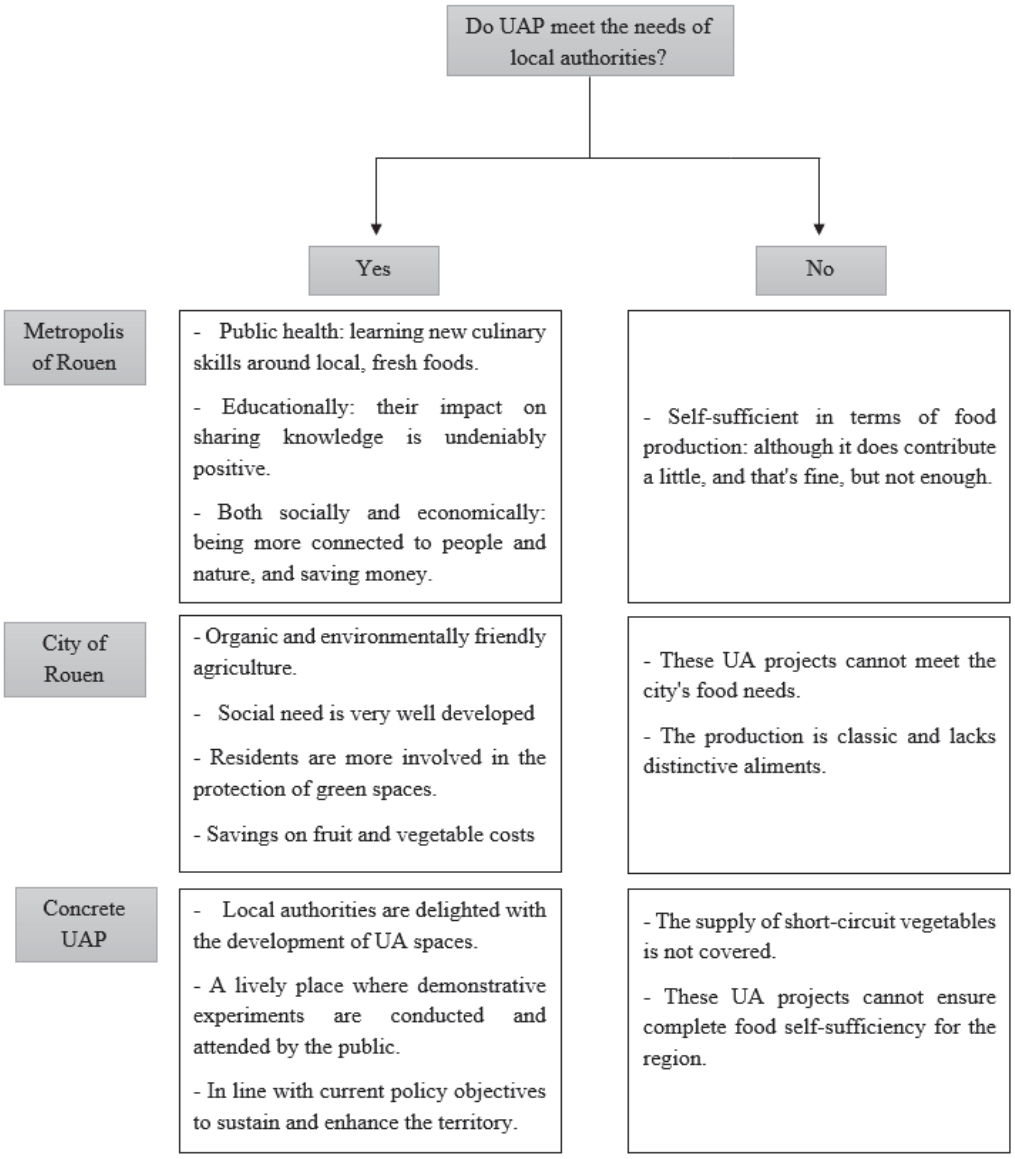


Figure 7. Evaluation of the effectiveness of the established links and the assessment of the extent of governmental interaction between the three entities.

4.7. Limitations and Potential Directions

This research revealed various limitations, focusing in particular on an aspect that is difficult to address in the short term: the city's food sufficiency and the long-term viability of UAP. Indeed, the findings from all three entities confirm that one of the main purposes of implementing such projects is to feed the city's inhabitants. However, this is a challenging objective given the region's high population and limited urban areas. As for long-term viability, local authorities have indicated that this is a major obstacle, as they are keen to contribute to the realization and implementation of the UAP but need to be sure that it will last for many years. On the presidents' side, these same limitations concern the urban areas available for the creation of such projects, since they complain that "it's difficult to find land on which to create an urban space given that today all the land is left for the construction of buildings".

Although the three entities recognize that significant improvements cannot be accomplished so quickly, they strongly believe that potential directions and future improvements can be achieved through the implementation of new forms of UA (vertical farming, rooftops, etc.) as well as through efficient water management and, in particular, through the implementation of future plans to integrate UA into city plans contributing to both sustainable food systems and city resilience.

5. Discussion

This research paper analyzes the interactions between different entities in a framework of MLG, focusing on the insertion of the urban sector within cities in order to analyze whether the expectations of local authorities are being realized by urban practitioners in the field. Conducted in the CR and its metropolitan area, this research is based on in-depth interviews with actors from each entity, including representatives of the MR, the CR, and presidents from two UAP, which yielded 28 interviews, a sufficient number to answer all our questions and solve our problematics. The methodological approach used in this article is well chosen to elicit the diversity of opinions on the various topics [30] and offers a valuable qualitative approach to the analysis of the various thematic and stakeholder contributions at different levels while also including quantitative results to make the analysis completer and more accurate.

Common features between the three entities were observed on the basis of the strategic objectives resulting from this investigation, where local entities emphasize that the creation of urban spaces ranges from the need to ensure healthy, cost-saving, and appropriate food to sharing new learning and connecting city dwellers to local environments, where this is perfectly achieved by the urban practices evaluated. These dimensions match those outlined by Menconi et al. [15] for the three pillars of sustainability: social, economic, and environmental, which are key objectives in the application of UAP. These urban projects aim not only to save money on a day-to-day basis but also to provide employment for retirees or job seekers. This finding has also been confirmed by [54,55], who reported significant potential savings after their involvement and employment in UAP.

The analysis also focuses on ensuring an integrated framework between the various entities to facilitate collective action among them. Indeed, the results obtained from the three entities underline the importance of an appropriate participatory approach through PG to facilitate fruitful partnerships and reconcile conflicting stakeholder interests by improving communication mechanisms and aligning objectives while ensuring combined efforts can have a greater influence on UA and make our cities more sustainable and healthier places to live. In the case of this research, this approach is positively affirmed by all stakeholders, since there is a reciprocity of satisfaction at different scales and levels. This perspective is also supported by Saint-Ges et al. [56], who assert that such an approach maximizes synergistic interactions between stakeholders while ensuring the resilient use of resources and infrastructure.

Apart from the common criteria mentioned above, in terms of food self-sufficiency, divergences remain, as the local authorities are keen to have independent and local cities and regions through the UAP, while urban practitioners do not and cannot meet the city's needs in terms of food supply due to the limited available area set aside for the UAP and the city's soil and pollution limitations. Moreover, our interlocutors at the MR and CR recognize these limitations but are nonetheless very satisfied with the current results that UAP is having on the residents and the city itself. However, local authorities still consider some indicators for the proper operation of the UAP, where the metropolis includes criteria relating to the number and integration of participants in projects, as well as the quantity of activities carried out per season and per year. The city, on the other hand, focuses more on setting standards for the surface area allocated to UA while imposing verification by urban planners to ensure the absence of future projects that could compromise the sustainability of UA initiatives. In this case, we need to consider the principle of subsidiarity, which means that responsibilities should be decentralized to the lowest level of governance that

can do the job and manage the task, something that was also mentioned by Marshall in 2008 [57] as a way of better understanding how responsibilities should be distributed between different levels of governance and that higher levels should not take on tasks that lower levels can manage.

The results reveal another important constraint related to the criteria imposed on UAP developers. These differences can be explained by the availability of land: at the metropolitan level, vacant spaces intended for the insertion of urban spaces are considerably more extensive than within the city. As a result, the latter must comply with strict surface area requirements and future planning regulations drawn up by urban planners and architects. These dynamics highlight the need to adapt the criteria for implementing UAP to the contextual specificities of each entity while considering space constraints and urban planning objectives. Nevertheless, UAP presidents and owners respect these requirements and would nevertheless appreciate greater flexibility.

6. Conclusions

UA is seen as a key lever for promoting sustainable and balanced nutrition in urban areas [20]. It also plays a role in social integration and well-being while generating economic and environmental benefits [26]. To achieve these objectives requires rigorous and sustainable management of PG in the implementation of these projects [20], and that is what this research article has investigated.

This investigation highlighted an essential aspect of the governance and institutional framework of UAP, particularly in the context of PG. Indeed, the results show that the MR and CR monitor the implementation of these projects and provide regular support to long-term project leaders. This support for collaborative working is underlined by the feedback from CD and JA project leaders, who express their satisfaction with working with local authorities, stating that they are "happy and satisfied to work with them". This underlines the strength of the organizational structure and the dynamic collaboration facilitated by the PG between local authorities and project leaders, providing an encouraging supportive environment for potential UAP owners to become more actively involved and committed to this field.

This research is one of the first to investigate the link between different entities in the UA field through a PG analysis and using an MLP approach, thus highlighting its original contribution. Indeed, this analysis assesses the results of three different entities separately, in order to reach a single comprehensive statement, which in this case is to analyze whether the expectations of local authorities are aligned with the reality of projects implemented in the CR, France. Future research might consider using this paper as a reference while following the same methodology of a combined approach between what is stated and mentioned by local authorities and what is actually implemented in different cities and countries, in order to have a complete picture of the actual situation and to determine in this case whether there is complementarity between the needs and main objectives of local authorities in implementing UAP and the real-life practices implemented. Furthermore, this document could be useful for planners, architects, and urban specialists to better understand, design, and shape the future dynamics of the UAP by promoting sustainable local food and resilient cities for the future, in line with the three aspects of sustainability, grouped into the economic, social, and environmental dimensions.

Author Contributions: Conceptualization, O.F.B., L.M.N.-G. and L.C.-S.; methodology, O.F.B., L.C.-S., L.S., L.M.N.-G. and F.P.-F.; software, O.F.B., L.S. and L.C.-S.; validation, L.C.-S., L.M.N.-G. and L.S.; formal analysis, O.F.B., L.S., L.C.-S. and L.M.N.-G.; investigation, O.F.B., F.P.-F. and L.C.-S.; resources, L.M.N.-G. and L.C.-S.; writing—original draft preparation, O.F.B.; writing—review and editing, O.F.B., L.C.-S., L.S., L.M.N.-G. and F.P.-F.; visualization, O.F.B., L.C.-S., L.M.N.-G. and L.S.; supervision, L.S., L.C.-S. and L.M.N.-G.; project administration, L.M.N.-G.; funding acquisition, L.M.N.-G. and L.C.-S. All authors have read and agreed to the published version of the manuscript.

Funding: The present work is related to the FUSILLI Project, which is funded by the European Union's Horizon 2020 Research and Innovation program under the grant agreement No. 101000717 (https://fusilli-project.eu/) (accessed on 10 December 2024).

Data Availability Statement: Data are contained within the article.

Acknowledgments: Ouiam Fatiha Boukharta has been financed under the call for University of Valladolid 2021 predoctoral contracts, co-financed by Banco Santander.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of this study, in the collection, analysis, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

CR	City of Rouen
CP	«Le champ des possibles»
JA	«Le jardin de l'Astéroïde»
MLG	Multi-level governance
MR	Metropolis of Rouen
PG	Polycentric governance
UA	Urban agriculture
UAP	Urban agricultural projects

References

1. Kolagar, M. Adherence to urban agriculture in order to reach sustainable cities; a BWM–WASPAS approach. *Smart Cities* **2019**, *2*, 31–45. [CrossRef]
2. Ahmed, S.; Byker Shanks, C. Supporting sustainable development goals through sustainable diets. In *Good Health and Well-Being*; Springer: Cham, Switzerland, 2020; pp. 688–699.
3. Mbow, C.; Rosenzweig, C.E.; Barioni, L.G.; Benton, T.G.; Herrero, M.; Krishnapillai, M.; Diouf, A.A. *Food Security*; No. GSFC-EDAA-TN78913; IPCC: Geneva, Switzerland, 2020.
4. Karthika, D.P.; Karthikeyan, P. Urbanization, food insecurity and agriculture–challenges for social sustainable development. *Probl. Ekorozwoju–Probl. Sustain. Dev.* **2016**, *12*, 157–162.
5. Knickel, K.; Zemeckis, R.; Tisenkopfs, T. A critical reflection of the meaning of agricultural modernization in a world of increasing demands and finite resources. In Proceedings of the International Scientific Conference «Rural development 2013», Akademija, Lithuania, 28–29 November 2013; Volume 6, pp. 561–567.
6. Boukharta, O.F.; Huang, I.Y.; Vickers, L.; Navas-Gracia, L.M.; Chico-Santamarta, L. Benefits of Non-Commercial Urban Agricultural Practices—A Systematic Literature Review. *Agronomy* **2024**, *14*, 234. [CrossRef]
7. Mougeot, L.J. *Urban Agriculture: Definition, Presence, Potentials and Risks, and Policy Challenges*; Cities Feeding People Series; Report 31; IDRC: Ottawa, ON, Canada, 2000.
8. Ferreira, A.J.; Guilherme, R.I.; Ferreira, C.S. Urban agriculture, a tool towards more resilient urban communities? *Curr. Opin. Environ. Sci. Health* **2018**, *5*, 93–97. [CrossRef]
9. Pojani, D.; Stead, D. Sustainable urban transport in the developing world: Beyond megacities. *Sustainability* **2015**, *7*, 7784–7805. [CrossRef]
10. Mohiuddin, M.; Al Mamun, A.; Syed, F.A.; Mehedi Masud, M.; Su, Z. Environmental knowledge, awareness, and business school students' intentions to purchase green vehicles in emerging countries. *Sustainability* **2018**, *10*, 1534. [CrossRef]
11. Azunre, G.A.; Amponsah, O.; Peprah, C.; Takyi, S.A.; Braimah, I. A review of the role of urban agriculture in the sustainable city discourse. *Cities* **2019**, *93*, 104–119. [CrossRef]
12. Frimpong Boamah, E. A polycentric vision for governing the urban food commons. In *Planning for Equitable Urban Agriculture in the United States: Future Directions for a New Ethic in City Building*; Springer International Publishing: Cham, Switzerland, 2024; pp. 121–141.
13. Aligica, P.D.; Tarko, V. Polycentricity: From Polanyi to Ostrom, and beyond. *Governance* **2012**, *25*, 237–262. [CrossRef]
14. Manganelli, A. *The Hybrid Governance of Urban Food Movements: Learning from Toronto and Brussels*; Springer Nature: Berlin/Heidelberg, Germany, 2022.
15. Menconi, M.E.; Heland, L.; Grohmann, D. Learning from the gardeners of the oldest community garden in Seattle: Resilience explained through ecosystem services analysis. *Urban For. Urban Green.* **2020**, *56*, 126878. [CrossRef]
16. Fleury, A.; Ba, A. Multifunctionality and sustainability of urban agriculture. *Urban Agric. Mag.* **2005**, *15*, 4–6.
17. Cohen, N. *Policy Brief: New Directions for UA in NYC*; CUNY Urban Food Policy Institute: Harlem, NY, USA, 2016.
18. Santo, R.; Palmer, A.; Kim, B. *Vacant Lots to Vibrant Plots: A Review of the Benefits and Limitations of Urban Agriculture*; Johns Hopkins Center for a Livable Future: Baltimore, MD, USA, 2016.

19. Horst, M.; McClintock, N.; Hoey, L. The intersection of planning, urban agriculture, and food justice: A review of the literature. In *Planning for Equitable Urban Agriculture in the United States: Future Directions for a New Ethic in City Building*; Springer: Cham, Switzerland, 2024; pp. 89–120.
20. Siegner, A.; Sowerwine, J.; Acey, C. Does urban agriculture improve food security? Examining the nexus of food access and distribution of urban produced foods in the United States: A systematic review. *Sustainability* **2018**, *10*, 2988. [CrossRef]
21. Mailhot-Leduc, I. L'agriculture Urbaine et la Mise en Débat du Projet Urbain: Une Consultation Publique à Montréal. Ph.D. Thesis, Université du Québec à Montréal, Montréal, QC, Canada, 2014.
22. De Oliveira, J.A.; Ahmed, A. Governance of urban agriculture in African cities: Gaps and opportunities for innovation in Accra, Ghana. *J. Clean. Prod.* **2021**, *312*, 127730. [CrossRef]
23. Boukharta, O.F.; Pena-Fabri, F.; Chico-Santamarta, L.; Navas-Gracia, L.M.; Sauvée, L. Governance structures and stakeholder's involvement in Urban Agricultural projects: An analysis of four case studies in France. *Int. Food Agribus. Manag. Rev.* **2023**, *27*, 76–93. [CrossRef]
24. Morrison, T.H. Evolving polycentric governance of the Great Barrier Reef. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E3013–E3021. [CrossRef]
25. Jordan, A.J.; Huitema, D.; Hildén, M.; Van Asselt, H.; Rayner, T.J.; Schoenefeld, J.J.; Tosun, J.; Forster, J.; Boasson, E.L. Emergence of polycentric climate governance and its future prospects. *Nat. Clim. Chang.* **2015**, *5*, 977–982. [CrossRef]
26. Morrison, T.H.; Bodin, Ö.; Cumming, G.S.; Lubell, M.; Seppelt, R.; Seppelt, T.; Weible, C.M. Building blocks of polycentric governance. *Policy Stud. J.* **2023**, *51*, 475–499. [CrossRef]
27. Lemos, M.C.; Agrawal, A. Environmental governance. *Annu. Rev. Environ. Resour.* **2006**, *31*, 297–325. [CrossRef]
28. Angst, M.; Mewhirter, J.; McLaughlin, D.; Fischer, M. Who joins a forum—And who does not?—Evaluating drivers of forum participation in polycentric governance systems. *Public Adm. Rev.* **2022**, *82*, 692–707. [CrossRef]
29. McGinnis, M.D. Networks of adjacent action situations in polycentric governance. *Policy Stud. J.* **2011**, *39*, 51–78. [CrossRef]
30. Balmer, D.F.; Richards, B.F. Conducting qualitative research through time: How might theory be useful in longitudinal qualitative research? *Adv. Health Sci. Educ.* **2022**, *27*, 277–288. [CrossRef]
31. Creswell, J.W. *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*; Pearson: London, UK, 2015.
32. Sandall, B.K. Investigating Educators' Perceptions of STEM Integration: A Semi-Structured Interview Approach. Ph.D. Thesis, University of Nebraska, Lincoln, NE, USA, 2016.
33. Robinson, O.C. Sampling in interview-based qualitative research: A theoretical and practical guide. *Qual. Res. Psychol.* **2014**, *11*, 25–41. [CrossRef]
34. Thurman, N. Mixed methods communication research: Combining qualitative and quantitative approaches in the study of online journalism. In *SAGE Research Methods Cases*; Sage: Thousand Oaks, CA, USA, 2018.
35. Clarke, V.; Braun, V. Using thematic analysis in counselling and psychotherapy research: A critical reflection. *Couns. Psychother. Res.* **2018**, *18*, 107–110. [CrossRef]
36. Turk, M.; Turk, S.T.; Muftuoglu, A.C.; Karakaya, O.; Karakaya, K. Students' Expectations and Experiences about Engagement Strategies in Online Courses: A Mixed Methods Study. *Online Learn.* **2024**, *28*, n2. [CrossRef]
37. Oliffe, J.L.; Kelly, M.T.; Gonzalez Montaner, G.; Yu Ko, W.F. Zoom interviews: Benefits and concessions. *Int. J. Qual. Methods* **2021**, *20*, 16094069211053522. [CrossRef]
38. Killawi, A.; Khidir, A.; Elnashar, M.; Abdelrahim, H.; Hammoud, M.; Elliott, H.; Thurston, M.; Asad, H.; Al-Khal, A.L.; Fetters, M.D. Procedures of recruiting, obtaining informed consent, and compensating research participants in Qatar: Findings from a qualitative investigation. *BMC Med. Ethics* **2014**, *15*, 9. [CrossRef]
39. Karvonen, A.; Van Heur, B. Urban laboratories: Experiments in reworking cities. *Int. J. Urban Reg. Res.* **2014**, *38*, 379–392. [CrossRef]
40. Suckert, L. Back to the future. sociological perspectives on expectations, aspirations and imagined futures. *Eur. J. Sociol./Arch. Eur. Sociol.* **2022**, *63*, 393–428. [CrossRef]
41. Nesbitt, L.; Hotte, N.; Barron, S.; Cowan, J.; Sheppard, S.R. The social and economic value of cultural ecosystem services provided by urban forests in North America: A review and suggestions for future research. *Urban For. Urban Green.* **2017**, *25*, 103–111. [CrossRef]
42. Pahl-Wostl, C. Governance of the water-energy-food security nexus: A multi-level coordination challenge. *Environ. Sci. Policy* **2019**, *92*, 356–367. [CrossRef]
43. Caughman, L.; Beaudoin, F.; Withycombe Keeler, L. The project-partnership cycle: Managing city-university partnerships for urban sustainability and resilience transformations. *Urban Transform.* **2023**, *5*, 10. [CrossRef]
44. Insee. Estimations de Population par Sexe et âge au 1er Janvier 2024. 2024. Available online: https://www.insee.fr/fr/statistiques/8069529 (accessed on 10 December 2024).
45. Vial, C.; Barget, E. *Les événements Sportifs et le Développement Durable: Le cas des Jeux Equestres Mondi-Aux FEI Alltech 2014 en Normandie*; Publibook/Société écrivains: Paris, France, 2018.
46. DRAFF. Atlas des Territoires Agricoles de la NorMandie. Available online: https://www.prefectures-regions.gouv.fr/normandie/content/download/56045/369480/file/20190128-DRAFF-Atlas-version-light.pdf (accessed on 10 December 2024).

47. Nouaceur, Z.; Murărescu, O.; Murătoreanu, G. Study of Air and Surface Temperatures in the City of Rouen (France) in Order to Evaluate the Characteristics of the Urban Heat Islands. Available online: https://georeview.usv.ro/wp-content/uploads/2023/07/Article.12-Vol.33-1.pdf (accessed on 10 December 2024).
48. Gómez-Álvarez, D.; Rajack, R.M.; López-Moreno, E.; Lanfranchi, G. (Eds.) *Steering the Metropolis: Metropolitan Governance for Sustainable Urban Development*; Inter-American Development Bank: Washington, DC, USA, 2017.
49. Lestel, L.; Meybeck, M.; Carré, C.; Belliard, J. The seine, the river dedicated to Paris. In *River Culture–Life as a Dance to the Rhythm of the Waters*; UNESCO Publishing: Paris, France, 2023; pp. 673–697.
50. Blay-Palmer, A.; Santini, G.; Dubbeling, M.; Renting, H.; Taguchi, M.; Giordano, T. Validating the city region food system approach: Enacting inclusive, transformational city region food systems. *Sustainability* **2018**, *10*, 1680. [CrossRef]
51. García-Llorente, M.; Rossignoli, C.M.; Di Iacovo, F.; Moruzzo, R. Social farming in the promotion of social-ecological sustainability in rural and periurban areas. *Sustainability* **2016**, *8*, 1238. [CrossRef]
52. Hall, C.M. The ecological and environmental significance of urban wastelands and drosscapes. In *Organising Waste in the City*; Policy Press: Bristol, UK, 2013; pp. 21–40.
53. Anguelovski, I.; Connolly, J.; Brand, A.L. From landscapes of utopia to the margins of the green urban life: For whom is the new green city? *City* **2018**, *22*, 417–436. [CrossRef]
54. Tulla, A.F.; Vera, A.; Valldeperas, N.; Guirado, C. Social return and economic viability of social farming in Catalonia: A case-study analysis. *Eur. Countrys.* **2018**, *10*, 398–428. [CrossRef]
55. McLarnon, M. Investigating the Social Relations of Community Gardens for Adult Education. *Can. J. Study Adult Educ.* **2024**, 36. [CrossRef]
56. Saint-Ges, V.; Warin, B.; Jacobsohn, A. Public Urban Agriculture Equipment: A Tool for Fair and Sustainable Food Policies? Potager du Roi and Cité Maraîchère, Two French Case Studies. *Sustainability* **2024**, *16*, 7399. [CrossRef]
57. Marshall, G.R. Nesting, subsidiarity, and community-based environmental governance beyond the local level. *Int. J. Commons* **2008**, *2*, 75–97. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

sustainability

MDPI

Article

Remaining Life Prediction of Automatic Fare Collection Systems from the Perspective of Sustainable Development: A Sparse and Weak Feature Fault Data-Based Approach

Jing Xiong [1,2,*], Youchao Sun [1], Zhihao Xu [3], Yongbing Wan [4] and Gang Yu [3]

[1] College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China; sunyc@nuaa.edu.cn
[2] College of Air Transportation, Shanghai University of Engineering Science, Shanghai 201620, China
[3] SILC Business School, Shanghai University, Shanghai 201800, China; xuz67872@gmail.com (Z.X.); gyu@shu.edu.cn (G.Y.)
[4] Shanghai Rail Transit Technology Research Center, Shanghai 201103, China; wybingsh@163.com
* Correspondence: xjsh@sues.edu.cn; Tel.: +86-021-67791109

Abstract: The most effective way to solve urban traffic congestion in mega cities is to develop rail transit, which is also an important strategy for sustainable urban development. Improving the service performance of rail transit equipment is the key to ensuring the sustainable operation of urban rail transit. Automatic fare collection (AFC) is an indispensable system in urban rail transit. AFC directly serves passengers, and its condition directly affects the sustainability and safety of urban rail transit. This study proposes remaining useful life (RUL) prediction framework for AFC systems. Firstly, it proposes the quantification of AFC health state based on health degree, and proposes a health state assessment method based on digital analog fusion, which compensates for the shortcomings of single data-driven or model driven health methods. Secondly, it constructs a multi feature extraction method based on multi-layer LSTM, which can capture long-term temporal dependencies and multi-dimensional feature, overcoming the limitation of low model accuracy because of the weak data features. Then, the SSA-XGBoost model for AFC RUL prediction is proposed, which effectively performs global and local searches, reduces the possibility of overfitting, and improves the accuracy of the prediction model. Finally, we put it into practice of the AFC system of Shanghai Metro Line 10. The experiment shows that the proposed model has an MSE of 0.00111 and MAE of 0.02869 on the test set, while on the validation set, MSE is 0.00004 and MAE is 0.00659. These indicators are significantly better than other comparative models such as XGBoost, random forest regression, and linear regression. In addition, the SSA-XGBoost model also performs well on R-squared, further verifying its effectiveness in prediction accuracy and model fitting.

Keywords: AFC system; remaining useful life prediction; sparse and weak feature data; health assessment; XGBoost

1. Introduction

The urban rail transit system has become an important means of relieving traffic congestion in major cities around the world. In 2023, China's rail transit operation mileage reached 9584 km, with 5609 stations, 57,286 assigned vehicles, and a total number of passenger trips of 2.37 billion. Facing an increasingly complex line network, increasing train frequency, growing passenger flow, and the continuous, uninterrupted operation of urban rail transit systems, the maintenance costs of urban rail transit equipment are rising.

The demand for intelligent operation and maintenance (O&M) is becoming increasingly prominent. A good urban rail transit equipment system can effectively ensure the sustainable development of urban rail transit. Therefore, intelligent O&M has become key to the sustainable operation development of rail transit.

Automatic fare collection (AFC) is an indispensable system in urban rail transit. AFC directly serves passengers, and its status directly affects the service level of urban rail transit. The AFC system includes multiple components from automatic ticket vending machines to central clearance rooms, and it is crucial to the sustainability and safety of urban rail transit operations [1–4]. The 2022 Wuhan Metro station-wide AFC equipment power-down fault resulted in all AFC equipment not being able to sell tickets, and AFC fan doors continued to open, affecting the smoothness of passenger entry and exit, causing passengers to be stranded, and seriously affecting the operational efficiency of the metro station and the travelling experience of the passengers. At present, China's urban rail transit AFC equipment maintenance strategy is mainly after-the-fact maintenance and regular maintenance. However, the after-action maintenance strategy, due to the timeliness of fault reporting, the accuracy of fault information, the timeliness of the response, and other factors, can easily lead not only to a decline in the quality of the equipment, but also lead to a decline in the level of service. Regular maintenance, due to inappropriate limits on overhaul cycles, means parts are prone to many faults before replacement or to excessive and unnecessary maintenance work, resulting in high maintenance costs and wasted maintenance resources. Changes in operation and maintenance mode remain a huge challenge for the sustainable operation and development of urban rail transit.

To meet the demand of the sustainable operation development of rail transit in China, research into intelligent O&M technology of rail transit station equipment and facilities is becoming more and more urgent. For AFC systems, this involves many components; each component has different fault modes and maintenance requirements, and each fault has a uniqueness and coupling relationship, so it is difficult to explore the system's fault mechanism, and this means that the physically based model has limitations in the RUL prediction of AFC systems. The noise, complexity, nonlinearity, and non-stationarity of fault data directly affect the accuracy and applicability of data-driven remaining-life-prediction methods. Therefore, their RUL prediction for sustainable operation still face great challenges [5,6].

To solve these challenges, this paper aims to develop a new RUL prediction method for AFC systems of urban rail transit. Firstly, to compensate for the scarcity and variable quality of raw data, this paper uses Weibull distribution and generative adversarial networks (GAN) to enhance data diversity and magnitude in model training. Secondly, a health assessment method based on digital analog fusion is designed to evaluate the equipment health state using historical fault data, environmental factors and expert knowledge, which optimizes the processing of complex decision-making problems and reduces the influence of the subjective judgement of experts in the empirical-based method. Then, this paper adopts a multi-layer long short-term memory (LSTM) network architecture to capture long-term dependencies and non-linear characteristics in data, and proposes a life prediction model based on SSA optimization, which efficiently performs both global and local searches, and reduces the likelihood of overfitting.

The paper is organized as follows: Section 2 introduces the related research about prognostics and health management. Section 3 details the proposed health status evaluation method based on digital analog fusion and the SSA-optimized XGBoost RUL prediction model. Section 4 presents the application and results analysis of the proposed method on Shanghai Metro Line 10. Section 5 provides the summary and future work.

2. Literature Review

Prognostics and health management (PHM) [7] is based on the design and validation of data and information acquisition of application objects, and through a series of data analysis techniques such as data information preprocessing, data classification, feature extraction and a series of data analysis techniques, obtains the required data forms, and finally, through the application of prediction algorithms and the integration of system performance indicators, it performs the object's health status assessment and remaining life prediction.

2.1. Data-Driven Health Assessment of Electromechanical Equipment

The main methods for the health state assessment of electromechanical equipment include knowledge-based experience [8,9], state model-based methods [10], and data-driven methods [11]. Zhang et al. present a novel health indicator for the health assessment of lithium-ion batteries [12]. Ref. [13] used CEEMDAN to reduce modal confounding effects and reconstruct variance features, and used the cumulative function (CF) to construct health values. Guo et al. designed a model for the health status assessment of electric vehicle batteries focusing on the polarization resistance and ohmic resistance of the batteries [14]. Lee et al. estimated and diagnosed the state-of-charge and state-of-health of a battery using multilayer neural networks (MNNs) [15].

To solve the real-time health assessment, Zhang et al. proposed the concept of health degree (HD), and proposed the NISVDD algorithm to achieve online updating of the data [16]. Zhou used SG-FCM clustering to generate different health status clusters and evaluate the health status using these clusters to improve the accuracy; the researchers jointly applied expert knowledge to improve the data-driven model [17]. Jiang used a Bayesian network to assess the health status of satellite batteries [18]. Mi established a motor system health assessment by dynamic fault tree and Bayesian network combined with domain expert experience [19].

2.2. Data-Driven Prediction of Remaining Useful Life

RUL prediction methods include physical model-based [20,21], data-driven [22–24] and hybrid model-based methods [25,26]. However, physical modelling approaches face challenges when dealing with big data and highly non-linear systems, requiring complex model calibration and extensive expertise. Data-based prediction methods do not require an understanding of the equipment degradation process and only use machine learning and deep learning algorithms to process and analyze the data to uncover patterns and trends inherent in equipment degradation to provide accurate RUL predictions. Chang used the LSTM algorithm and optimized it by grid search [27]. Other authors proposed a hybrid prediction model incorporating a convolutional neural network (CNN) and a stacked bidirectional unidirectional LSTM (SBU-LSTM) network [28]. In the field of bearing RUL prediction, attention mechanisms have gained popularity due to their ability to focus on the most relevant features. A dual-attention CNN method was proposed by Wang et al., combining a one-dimensional CNN and gated recurrent neural network (GRU) with Monte Carlo dropout simulation. This hybrid model provided better feature learning and uncertainty handling for lithium-ion battery degradation, demonstrating its potential for complex electromechanical systems like AFC [29]. Chen proposed prediction framework to efficiently deal with and utilize incomplete data by filling in the missing data through the GAIN model and performing RUL prediction using the MSDCNN-LSTM model [30]. Mo et al.'s study demonstrated a neural architecture search (NAS) technique using agent-assisted genetic algorithms to automatically discover the optimal architecture of a converter for RUL prediction [31]. Youness and Aalah developed an approach using explainable AI

with LSTM to enhance the interpretability of RUL predictions by restructuring features into homogeneous clusters. This method improved both learning efficiency and interpretability, allowing clearer insights into the model's decision-making process [32].

Hybrid prediction methods can combine the a priori knowledge of physical models with the flexibility of data-driven models. Chao et al. present an innovative hybrid framework for prediction of complex safety-critical systems [33]. Deng et al. combined the five-degrees-of-freedom dynamic model with particle filter-based calibration and proposed a hybrid calibration-based transfer learning framework for the cross-machine prediction of rolling bearing RUL [34]. Ref. [35] proposed a transformer method for prediction RUL of turbofan engines. Wang and Mamo proposed a hybrid method incorporating EWMA control charts and machine learning to predict the RUL of ball bearings [36]. Liao and Kottig developed hybrid prediction method by combining a classical Bayesian model with two data-driven methods [37]. Several advanced methodologies have been proposed for the prediction of remaining useful life in railway systems, particularly for electromechanical devices. A notable contribution in this area is the M2BIST-SPNet framework developed by Hu [38], which focuses on predicting the RUL of railway signaling electromechanical devices (RSEDs). The model employs a multi-scale multifrequency branch interactive spatiotemporal sequence prediction network (M2BIST-SPNet) that integrates several advanced mechanisms, including spatial-temporal attention-based convolution networks and communication modules, to effectively capture long-term multiparameter degradation trends.

Besides these above researches, ergonomic risk factors may also impact system reliability and remaining useful life. Mudiyanselage used sEMG wearable sensors and machine learning to assess automated workers' ergonomic risk in manual material handling [39]. Thus, the prediction of remaining useful life should also consider this factor.

2.3. Knowledge Gap

According to existing research, several challenges remain:

1. When faced with data scarcity or low-quality data, the performance of data-driven methods may degrade dramatically. In addition, data-driven models usually lack in-depth understanding of physical processes, which limits their interpretability and accuracy for application in complex systems.
2. Hybrid prediction models face high complexity and computational costs. Meanwhile, the development of hybrid prediction models usually requires extensive debugging and validation, which further increases the difficulty of model implementation.
3. Existing studies often lack effective treatment of unexpected events and nonlinear degradation processes in equipment operation. Many models fail to adequately account for dynamic changes and uncertainties in real application environments, which limits their application to real-time prediction and decision support.

These limitations become even more significant when considering the broader impacts on community connectivity. For instance, Halimi et al. illustrated how transportation infrastructure can either promote or disrupt community access, particularly in urban areas where connectivity challenges are more pronounced [40]. Such findings underscore the necessity for predictive models in AFC systems that are not only technically accurate but also resilient to prevent service interruptions that might otherwise isolate communities or reduce accessibility.

3. Methodology

3.1. Remaining Life Prediction Framework of AFC

To solve these above challenges, we collected operation and maintenance data for the AFC system of Shanghai Metro, which contained both structured and unstructured formats. Structured data include measurable parameters like usage statistics and fault counts, while unstructured data consist of maintenance reports, technician notes, and incident descriptions, which capture critical contextual information about system failures, environmental conditions, and operational history. However, due to the heterogeneity between structured and unstructured data, end-to-end models struggle with unstructured formats [41,42]. Thus, researchers usually use a hybrid or multi-modal approach to fully capture the data's heterogeneity and construct specific prediction models. Swaminathan et al. present a selective prediction model that improves the abstraction of unstructured clinical data, as end-to-end models fail to address these complexities and heterogeneity [43].

Similarly, we also need address problems caused by the nature of data, which are their heterogeneity, and sparse and weak features. The scarcity of high-quality data presents significant difficulties for data-driven models, which rely on enough diverse examples to accurately capture patterns of system degradation. Additionally, weak features make it harder to distinguish between critical and non-critical faults, limiting the model's ability to predict when a system component is likely to fail. We propose a prediction framework as shown in Figure 1, which mainly includes three parts: electromechanical equipment health assessment based on digital analog fusion, fault feature extraction based on multilayer LSTM, and prediction driven by SSA-XGBOOST.

In the stage of electromechanical equipment health assessment based on digital analog fusion, raw data are firstly subjected to data cleaning and normalization to improve the integrity and quality of the original data. Meanwhile, data augmentation methods based on Weibull distribution and GAN are used to generate enhanced data that conform to the characteristics of the raw data, which enhances data diversity and magnitude in model training. Then, evaluation indexes are established from the aspects of reliability, repairability, fault persistence, and fault severity, and health metrics are proposed to evaluate the health status of electromechanical equipment. A combination of the hierarchical analysis method (AHP) and the informativeness-based critic entropy method is used to compute the weights, which reduces the influence of subjective judgement of experts in the empirical-based method and achieves data-driven health metrics. To combine the equipment degradation mechanism, an accelerated degradation model (ADM)-based health metrication method is established on this basis. The ADM was selected due to its ability to simulate the physical degradation process of AFC systems under varying operational stress conditions, such as temperature fluctuations and overuse. The fusion of these two approaches is achieved by linearly combining their outputs, where the physical model provides a baseline understanding of the degradation process, and the data-driven model enhances the accuracy by refining weak data features and accounting for unexpected trends [44].

Another major challenge is the irregularity of fault occurrence times, which complicates the model's ability to learn long-term dependencies. These time-dependent relationships are critical for predicting the remaining life of system, but their complexity often leads to lower prediction accuracy. To address these issues, this paper proposes using advanced data augmentation and feature extraction methods to enrich the dataset and capture hidden temporal patterns, thereby improving the RUL prediction. In the stage of fault feature extraction based on multilayer LSTM, firstly, in order to capture long-term dependency and nonlinear characteristics, multilayer LSTM is established, and after extracting the environmental features and fault features of the original dataset by using a two-layer LSTM network, the two features are integrated to extract the deep-level healthiness features

affecting the equipment life through a third-layer LSTM network, so that the weak features existing in the original dataset can be solved. This solves the weak feature problem of the original dataset.

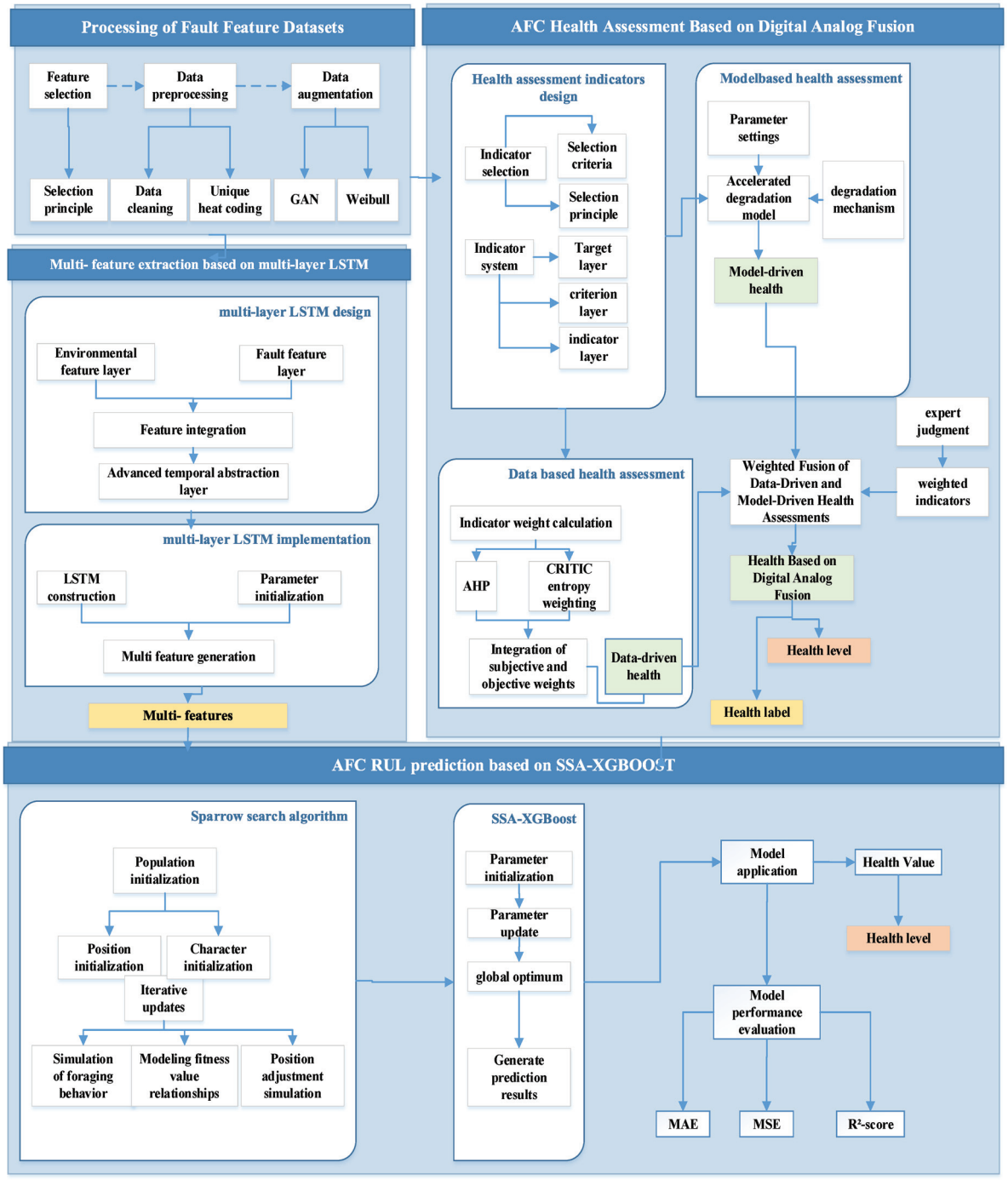


Figure 1. A framework for prediction of the remaining life of AFC.

The RUL prediction stage of electromechanical equipment based on SSA-XGBOOST uses the SSA algorithm to simulate sparrow's foraging behavior, adaptation and position adjustment to achieve the iterative updating process, which effectively carries out the global and local searches, optimizes the hyper-parameters of the XGBoost model, and solves the problems of under-convergence and overfitting due to the stochastic strategy of the XGBoost model.

3.2. AFC Health Assessment Based on Digital Analog Fusion

3.2.1. AFC Health Assessment Indicators and Weight Calculation

This paper establishes an assessment index by refining and quantifying the relative importance of each indicator and its sub-indicators, as shown in Table 1.

Table 1. Indicators for assessing health degree.

Goal Level	Guideline Level	Primary Indicator Level	Secondary Indicator Level
Health degree	Reliability	Relative importance of parts	Storage Host Network device Encoder
		Part fault type	Storage Mainframe Network device Encoder
	Repairability	Relative importance of spare parts	Battery Hard disk Capacitor Memory NIC Mainboard Controller
Health degree		Type of used spare parts	Battery Hard disk Capacitor Memory NIC Mainboard Controller
	Fault Persistence	Fault impact time	Less than 30 min Greater than 30 min
	Fault Severity	Fault grade	Primary fault Secondary fault Tertiary fault Other faults

Among them, the health state of AFC equipment is quantified using health level. Reliability considers the fault rate and the importance of the component, which indicate the ability of the equipment completes the specified functions under certain conditions and within a fixed period. Repairability uses the spare parts usage rate to indicate the ability of the equipment to maintain or return to normal operation after repair. Fault persistence uses fault impact time to indicate how difficult it is to return equipment to normal after a fault, usually in direct proportion. Fault severity uses the fault grade to describe the magnitude of the impact caused by an equipment fault.

In this paper, the four criteria will be reflected in the subjective weighting calculation using four level 1 indicators: relative importance of parts, relative importance of spares, time of fault impact and fault grade; and in the objective weighting calculation using the type of parts fault, type of spares used, time of fault impact and fault grade.

The fault impact time t_r characterizes the time required for the equipment to return to normal operation after a fault. The formula is:

$$t_r = t_1 - t_0 \tag{1}$$

where t_0 is the time to report the fault, t_1 is the time to restore the system, and t_r is zero if there is no impact on system operation.

Fault grades are used to characterize the extent of the impact caused by equipment faults and are obtained directly from the fault logging dataset.

3.2.2. Calculation of Weights for Subjective and Objective Evaluation Indicators

To determine the weights of each indicator proposed in Section 3.2.1, this paper uses a combination of subjective and objective weight confirmation methods. The AHP method was first used to calculate the subjective weights. An expert judgement matrix C is set, where C_{ij} denotes the value of importance of the ith criterion compared to the jth criterion.

$$C = \begin{bmatrix} 1 & C_{12} & \cdots & C_{1n} \\ \frac{1}{C_{12}} & 1 & \cdots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{C_{1n}} & \frac{1}{C_{2n}} & \cdots & 1 \end{bmatrix} \tag{2}$$

Secondly, by comparing the consistency test of the matrices, the weight vectors (W), and the consistency ratio (CR), the reasonableness of the judgement matrices is assessed. The closer the consistency ratio is to 0, the more consistent the matrix is. The method of calculation CR is as follows:

$$CR = \frac{\frac{\lambda_{max} - n}{n-1}}{RI} \tag{3}$$

where RI is a consistency index, λ_{max} is the maximum eigenvalue, n is the matrix order, and is the consistency index. To ensure that the sum of the weight vectors is 1, the weight vectors are normalized.

$$W_{normalized} = \frac{W}{\sum_{i=1}^{n} W_i} \tag{4}$$

The normalized weights of the sub-criteria are then used as their relative importance weights in the overall objective to form the final subjective indicator weights as follows:

$$\left\{ W_{normalized_fault_location}, W_{normalized_spare_utilization}, W_{normalized_system_recovery}, W_{normalized_fault_level} \right\} \tag{5}$$

To solve the limitation of single method of assigning weights, further critic method combined with entropy method is used for the calculation of objective weights. The assignment of weights in critic approach is calculated as follows:

$$\mathcal{W}_j = \frac{\sigma_j \sum_{i=1}^{n} (1 - r_{ij})}{\sum_{j=1}^{m} \sigma_j \sum_{i=1}^{n} (1 - r_{ij})} \tag{6}$$

where $\mathcal{W}_j$ is the jth indicator weight; r_{ij} is the correlation coefficient between indicators i and j, and σ_j is the standard deviation of indicator j.

In the critic entropy method, the uncertainty of the indicator values is measured by calculating the information entropy of each indicator, to derive the relative contribution of each indicator and then determine the weights. Firstly, for the secondary indicators in the part fault type, fault grade, spare parts usage type and system recovery time their fre-

quency of occurrence in the dataset is calculated. The frequency of occurrence is calculated as follows:

$$f_i = \frac{n_i}{N} \tag{7}$$

where f_i denotes the frequency of occurrence of secondary indicator i, n_i denotes the number of occurrences of secondary indicator i, and N denotes the total number of fault samples.

For the frequency of occurrence of each secondary indicator, its information entropy is calculated. The information entropy can be expressed by the following formula:

$$H = -\sum_{i=1}^{k} p_i \log_2 p_i \tag{8}$$

where p_i denotes the probability of occurrence of category i and k the total number of categories.

Finally, the objective weights of the indicators were calculated according to the following formula:

$$\mathcal{W}_j = \frac{(\sigma_j + e_j)\sum_{i=1}^{n}(1 - q_{ij})}{\sum_{j=1}^{m}(\sigma_j + e_j)\sum_{i=1}^{n}(1 - q_{ij})} \tag{9}$$

where $\mathcal{W}_j$ is the jth indicator weight, q_{ij} is the correlation coefficient between indicator i and j, σ_j is the standard deviation of indicator j, and e_j is the entropy of indicator j.

3.2.3. AFC Health Assessment

Considering the structural complexity and functional coupling of AFC equipment and the influence of physical wear and external impacts during operation, this paper proposes a digital analog fusion health degree quantitative calculation method to better reflect the trend of equipment health degree. Firstly, the calculation of health degree is carried out by combining the weights of subjective and objective indicators:

$$H_{original} = 1 - \sum_{i=1}^{n}\sum_{j=1}^{m} w_i w_j \tag{10}$$

To fully explore the relationship between fault data and time in the original dataset, this paper chooses the exponential degradation model to deal with the health degradation data of electromechanical equipment. According to the assumption of the exponential degradation model, the degradation process of the equipment can be expressed as a function of time, and the decay process of the health degree can be described by the exponential decay model in the ADM [45,46] framework in the following form:

$$H(t) = H_0 \times e^{-rt} \tag{11}$$

where $H(t)$ denotes the health degree at time t. H_0 represents the initial health degree and r denotes the health decay rate, as determined by expert experience.

Finally, a weighted average method was used to combine the original health degree and the health degree adjusted for time decay, as shown below:

$$H_{weighted}(t) = w_1 \times H_{original}(t) + w_2 \times H(t) \tag{12}$$

where $H_{weighted}$ represents the weighted health degree, w_1 and w_2 represent the weights of data health degree and model health degree, respectively, which are determined by expert empirical judgement. To further help the operator to understand and formulate maintenance strategies, this paper, based on Shanghai metro's standards for assessing the health state of operating facilities and equipment, as well as experts' experiences and related studies, classifies the health degrees obtained from the above studies into four

equipment health levels: A, B, C, and D. The way of dividing the health levels is shown in Table 2.

Table 2. Classification of health levels.

Level Name	Range	Description
Level A	0.80–1.00	Equipment is in good condition and maintenance intervals can be extended appropriately
Level B	0.65–0.80	Equipment needs to be monitored and regularly maintained
Level C	0.50–0.65	The equipment needs to strengthen the inspection and shorten the maintenance cycle
Level D	0.00–0.50	Equipment in urgent need of maintenance or replacement

3.3. Multi Feature Extraction Based on Multi-Layer LSTM

3.3.1. Data Augmentation Based on Weibull and GAN

To solve the data sparsity and imbalance problems of original data, this paper combines Weibull distribution and GAN for data augmentation. The Weibull distribution can simulate statistical characteristics of AFC equipment fault times, thus enhancing the accuracy of the model's prediction of fault occurrence times. GAN is used to generate fault pattern data that match the characteristics of the real data, which, in this case, extends the feature space of the dataset and increases the diversity and number of samples.

Firstly, historical fault time data are fitted using Weibull distribution to simulate future fault intervals based on the known fault history to provide data augmentation of the prediction model in the time dimension. The new fault time intervals generated through the Weibull distribution are shown below:

$$T_i = Weibull(\alpha, \beta) \tag{13}$$

where T_i, α, β denote the ith time interval of generation, the shape parameter and the scale parameter of the Weibull distribution, respectively.

The generator using the GAN then learns the distribution of data for generating new data, while the discriminator is used to identify generated data and real data. During the training process, these two parts compete and eventually the generator can produce increasingly realistic data. The generated data can be represented by the GAN model as:

$$X_{fake} = Generator(z) \tag{14}$$

where X_{fake} denotes the generated data, z is the random noise vector input to the generator.

3.3.2. Multidimensional Feature Extraction

Traditional time series analyses of AFC system faults tend to focus on single dimensional data, such as temperature or humidity. Several studies have demonstrated that LSTM networks, while effective in many time series applications, struggle to adequately capture certain complex, nonlinear characteristics inherent in sparse and weak feature data, particularly in scenarios where the data is irregular or lacks sufficient temporal patterns. To address these limitations, this paper employs a multi-layer LSTM architecture specifically designed to extract long-term dependencies and nonlinear relationships from multidimensional data in the context of the AFC system. To effectively obtain data features related to changes in the AFC system health status, this paper used multilayer LSTM to extract the feature inga as shown in Figure 2. Multiple dimensions of data, such as environmental monitoring data (temperature, humidity) and operation and maintenance data (fault type, location, etc.), are fused through the LSTM network to extract the time-dependent information at different levels, to capture the small changes of the system state more accurately. The output of each LSTM layer not only serves as an input for the next layer, but also

combines the memory states of previous layers, enabling the model to accumulate and transfer time-dependent features in the deep network structure. This fusion approach helps to reveal the interactions and influences between different data dimensions.

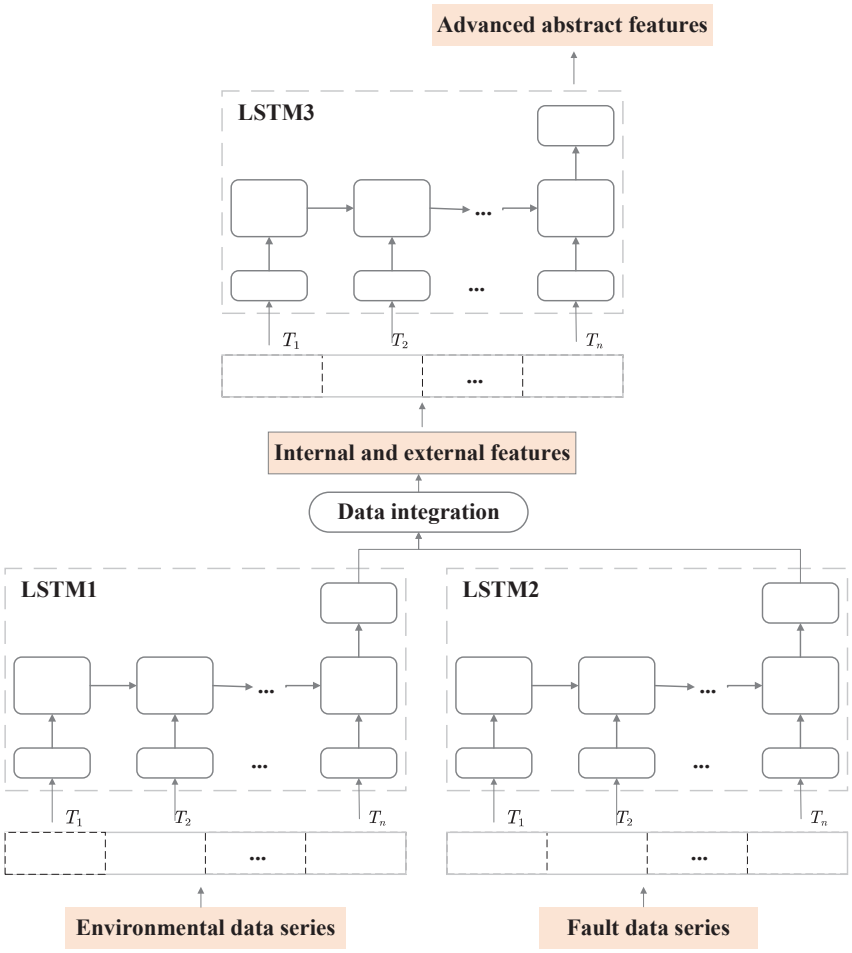


Figure 2. Multi-layer LSTM based feature extraction.

The multilayer LSTM shown in Figure 2 begins with a first LSTM layer constructed using a time series generator designed to capture time dependencies and patterns in the environmental data. This layer is configured with 64 LSTM units to ensure sufficient capacity to learn complex time series features and output feature representations through a Dense layer with 32 units.

The second layer input consists of the system's operational parameter variables such as fault location, fault level characterization, spare parts usage and system recovery time. These category-based variables are converted to numerical data after a solo thermal coding process. The second layer of the LSTM model also contains 64 cells and is output through a 32-cell dense layer.

After the first two layers of feature extraction, the input to the third LSTM layer consists of the output features from the first two layers, which are used to reveal deeper data associations and patterns. The third layer employs two LSTM layers configured with 512 units, where the first layer is set to return to the state of all time steps to retain the complete time series information, while the second layer further extracts the key features of these sequences. Finally, a 64-unit dense layer is used to output the refined resulting feature dataset.

3.4. AFC Remaining Life Prediction Based on SSA-XGBoost

3.4.1. XGboost

As shown in Figure 3, XGBoost constructs a model by combining multiple decision trees which form a tree structure including root, internal and leaf nodes for making decisions and generating possible outcomes. The model of XGboost algorithm can be represented as:

$$\hat{y}_i = f(x_i) = \sum_{k=1}^{k} f(x_i), f_k \in F \tag{15}$$

where $\hat{y}_i$ is the prediction of the ith sample; x_i denotes the ith input sample, K is the number of regression trees, f_k is the kth regression tree, and F denotes the set of K tree models.

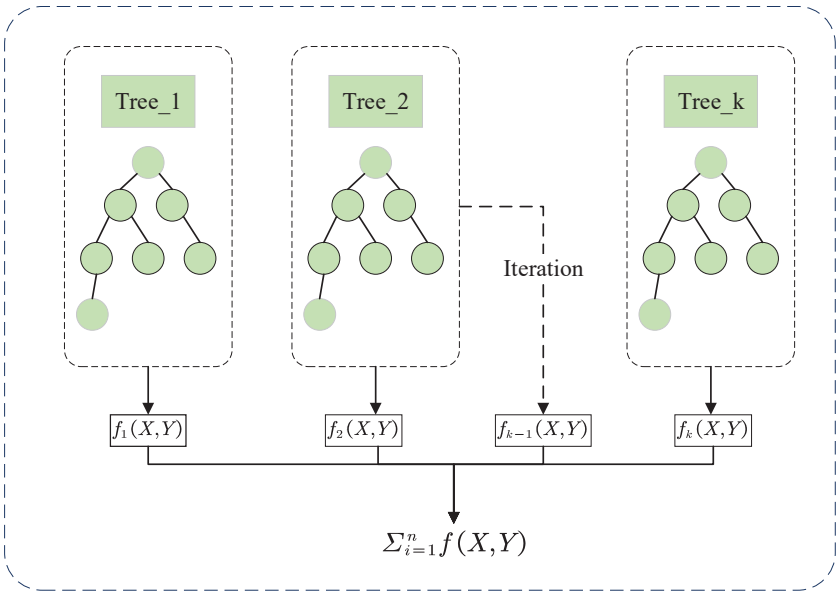


Figure 3. Structure of XGboost model.

In the model structure of XGBoost, each tree focuses on the residuals of the previous tree, branches from the root node, and continuously divides according to the data features to finally generate leaf nodes. These decision trees use a gradient algorithm to discover new ways of building trees to reduce residuals during model training. The algorithm branches out from the root node and finds the best split point for each branch based on the gradient information, so that each decision tree gradually fits the residuals of the data and continuously improves the prediction performance.

3.4.2. SSA-XGBoost

The XGBoost model usually has a large number of hyperparameters and high-dimensional search space. In order to effectively perform a global search in complex hyperparameter space, and to avoid falling into the local optimal solution, this paper uses the SSA algorithm to optimize XGBoost, which considers the hyperparameters of the XGBoost model as the solutions in the search space, through combining the search process of the SSA with the hyperparameter optimization of the XGBoost model, and using the simulation of cooperative work and information transfer in the sparrow population to dynamically adjust the key parameters of XGBoost and discover a better combination of parameters, and to find out more optimal combinations of parameters. By combining the search process of SSA with the optimization of hyperparameters of the XGBoost model, the key parameters of XGBoost are dynamically adjusted using the simulation of cooperative work and information transfer in the sparrow population. The steps of the SSA algorithm are as follows:

(1) The roles of discoverer and joiner:

Joiners are individuals with low fitness values and relatively insufficient energy reserves. They search for food according to the guidance of the discoverer, or forage around the discoverer. One individual becoming a discoverer inevitably leads to another individual becoming an accession to maintain the stability of the population. A population of dimension d consisting of n sparrows can be represented as:

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,d} \\ x_{2,1} & x_{2,2} & \dots & x_{2,d} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n,1} & x_{n,2} & \dots & x_{n,d} \end{bmatrix} \tag{16}$$

Whereas the fitness of sparrows can be expressed as:

$$F_x = \begin{bmatrix} f(x_{1,1}\ x_{1,2}\ \dots x_{1,d}\ ) \\ f(x_{2,1}\ x_{2,2}\ \dots x_{2,d}\ ) \\ \vdots \\ \vdots \\ f(x_{n,1}\ x_{n,2}\ \dots x_{n,d}\ ) \end{bmatrix} \tag{17}$$

(2) Alarm and search for safe areas.

When a foraging sparrow senses a predator's presence, it chirps as an alarm signal. When it is greater than the safety value, the foraging sparrow will guide the forager away from the dangerous area and forage in another safe area. This illustrates the collective behavior of sparrows in dangerous situations. The position update action can be expressed as follows:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^{t+1} \cdot exp\left(-\frac{i}{\alpha \cdot iter_{max}}\right)\ if\ R_2 < ST \\ X_{i,j}^{t+1} + Q \cdot L\ \ if\ R_2 \geq ST \end{cases} \tag{18}$$

(3) Relationship between energy reserves and fitness values.

The level of energy reserves of individual sparrows was related to their fitness values. Individuals with high energy reserves are more likely to be discoverers and to be able to persistently search for food-rich areas. Accessions with lower energy reserves may be limited by their energy level while foraging and are more likely to fly to other areas for more food.

(4) Joiner strategies.

Joiners are always able to search for the finder that provides the best food and either acquire food from it or search around it. To increase their chances of obtaining food, some joiners may compete for food resources and constantly monitor the location of finders to find and obtain food more quickly. The strategies of joiners may be expressed as follows:

$$X_{i,j}^{t+1} = \begin{cases} Q \cdot exp\left(\frac{X_{worst} - X_{i,j}^{t}}{\alpha \cdot iter_{max}}\right)\ if\ i > n/2 \\ X_p^{t+1} + \left|X_{i,j}^{t} - X_p^{t+1}\right| \cdot \mathrm{A}\ \text{otherwise} \end{cases} \tag{19}$$

(5) Hazard awareness and positional adjustment.

When aware of the presence of danger, sparrows at the edge of the colony moved quickly towards the safe area. Sparrows at the center of the group would walk randomly, trying to get closer to other sparrows. This behavioral pattern is simulated in an algorithm

to adjust the position of the sparrows and increase the population's ability to explore in the search space. Sparrows' danger-aware behavior can be expressed as:

$$X_{i,j}^{t+1} = \begin{cases} X_{bset}^{t} + \beta \cdot \left| X_{i,j}^{t} - X_{best}^{t} \right| \; if \; f_i > f_g \\ X_{i,j}^{t} + \left(\frac{\left| X_{i,j}^{t} - X_{worst}^{t} \right|}{\varepsilon + (f_i - f_w)} \right) \cdot \mathrm{K} \; if \; f_i = f_g \end{cases} \tag{20}$$

where X_{best}^{t} is the current global optimal position. β, as a step control parameter, is a normally distributed random number with mean 0 and variance 1. $\mathrm{K} \in [-1, 1]$ is a random number. f_i is then the fitness value of the current sparrow individual. f_g and f_w are the current global best and worst fitness values, respectively. ε is the smallest constant to avoid zeros in the denominator.

The optimized XGBoost operation mechanism based on SSA algorithm is shown in Figure 4.

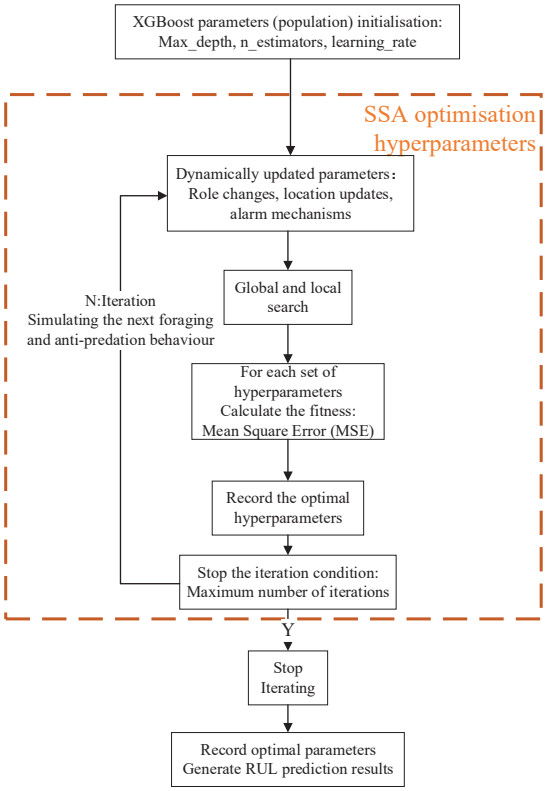


Figure 4. SSA-XGBoost based prediction process.

(1) Population initialization: Firstly, a population of solutions is defined, each solution represents a set of XGBoost hyperparameters, in this paper, *max_depth*, *n_estimators* and *learning_rate* are used. The parameter values for each individual in the population are randomly generated within the definition domain of the hyperparameters.

(2) Fitness function: A fitness function is defined, usually as a function of prediction error; this paper uses the MSE of the test set. The fitness of each individual is evaluated based on how its parameters affect the performance of the XGBoost model on a given dataset.

(3) Dynamic updating rules: the behavioral rules in the SSA algorithm, such as the roles of discoverers and joiners, the alarm mechanism, and the updating strategy of food location, are used to dynamically update the location of each individual in the population.

(4) Global and local search: SSA balances between global and local search, avoiding falling into local optimums by simulating the sparrow's foraging and vigilance behaviors, enhancing the algorithm's exploratory ability and the diversity of the solution space. This structured hyperparameter tuning process aligns with findings in [47], which emphasizes that comprehensive tuning across a defined hyperparameter space is essential for predictive accuracy in complex models.

(5) Stopping condition: the algorithm repeats the iteration until the termination condition is met, and this paper sets it to stop after reaching the maximum number of iterations.

4. Case Study

This paper uses the ACC-related dataset in the AFC system of Shanghai metro line 10 for model training and testing. This paper collects two parts of key data: equipment fault records and environmental monitoring data. The fault data came from the equipment fault records of the command center, which contain the fault events from January 2020 to June 2022. Each record details the exact time of the fault, type of fault, system and site involved, fault level, and system recovery time. The types of faults cover a wide range of dimensions, from encoder faults in the card production equipment to storage and mainframe faults in the master control center (MCC). Fault levels are categorized from level 1 to level 3, indicating the severity of the impact of the fault on system operations. The system recovery time provides the time span from the discovery of the fault to the resumption of normal operation, indicating the difficulty of repairing the fault. The environmental monitoring data were taken from the monthly inspection sheets of the ticket management department of the Shanghai railway operation and management center. The data include the temperature and humidity monitoring values recorded by the central host of AFC, in which the normal range of temperature is set at 18–25 °C and normal range of humidity is set at 20–80%, which are recorded in the form of numerical values directly.

Based on the above methodology, an application and validation method are proposed as shown in Figure 5, including four stages: data preprocessing, health state assessment, multi-feature extraction and RUL prediction.

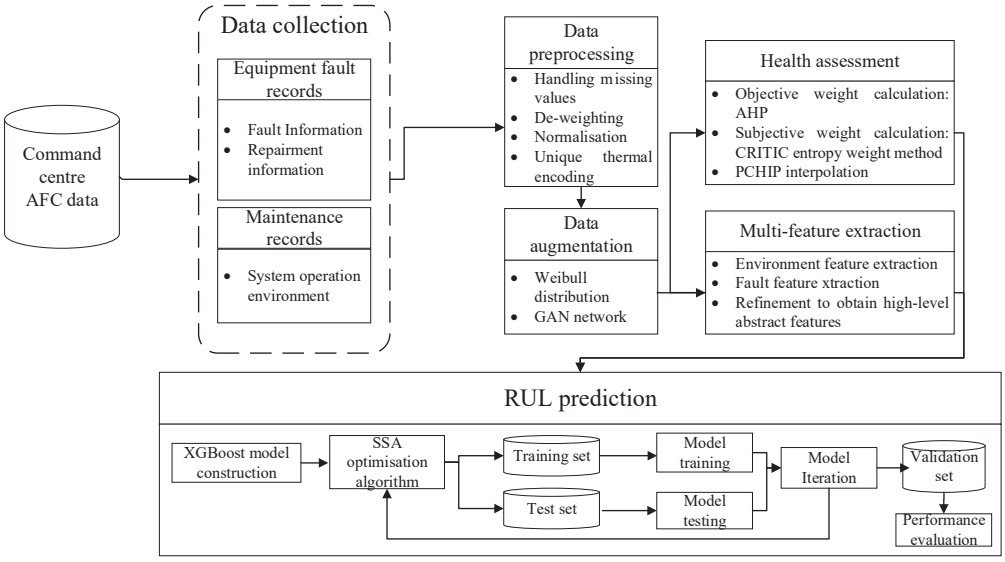


Figure 5. AFC system health state assessment and RUL prediction flowchart.

4.1. Data Preprocessing and Argument

This paper firstly selects the key features from the original dataset, including the environmental features such as temperature and humidity data, and the fault features

such as fault location and fault level characterization. Subsequently, the missing values were processed by filling in the missing values or deleting the records containing missing values. At the same time, the temperature and humidity data were normalized to match the requirements of the machine learning model inputs. Unique heat coding was performed on the categorical variables to transform the non-numerical data into a numerical form that could be processed by the machine learning model.

These preprocessing and augmentation methods not only enhance the model's technical accuracy but also contribute to a more equitable and accessible transportation system. Halimi et al. emphasizes the importance of data quality and accessibility in achieving equitable transportation solutions [48]. This aligns with our goal of ensuring that the model can support equitable and reliable AFC system operations, as reliable prediction and improved data handling reduce the risk of service interruptions that could impact accessibility for diverse urban populations.

Figure 6 shows the effect of data augmentation, with the blue curve indicating true loss of the discriminator, orange curve represents the hypothetical loss of the discriminator, and green curve represents the loss of the generator. At the beginning of training, the discriminator's ability to discriminate between real and synthetic data increases rapidly, as shown by the decrease in discriminator loss. As the training progresses, the generator starts to produce more and more realistic data, causing the discriminator loss to increase and the generator loss to decrease, showing a continuous improvement in the quality of the data it produces.

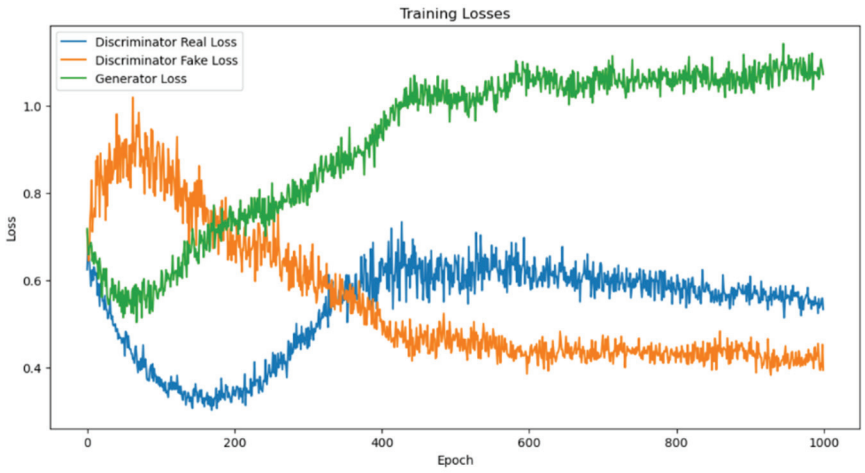


Figure 6. Loss curve of GAN network.

GANs involve a dynamic training process between the generator and the discriminator. As the training progresses, the discriminator becomes progressively better at distinguishing between real and generated data, which can result in the generator facing increasing difficulty in producing convincing samples. This leads to a phenomenon where the generator's loss may increase in the later stages of training. Yang et al. argue that this is not necessarily indicative of the generator failing but rather a reflection of the evolving nature of GAN training, where the discriminator's improvement makes it harder for the generator to "fool" the discriminator effectively [49]. After training, the loss curve tends to be stable, indicating that the GAN training has reached a stable state, which is a sign that the quality and diversity of the synthetic data meet the requirements of the f model. On this basis, the f data generated by Weibull distribution are combined with the fault feature data generated by GAN to form a complete fault instance to supplement the original dataset. Weibull distribution and GANs were chosen for data augmentation due to their ability to handle the complex, sparse, and weak-featured fault data in AFC systems. Weibull's flexibility in

modeling varying failure rates makes it suitable for generating realistic failure intervals, while GANs effectively generate synthetic data that preserves the underlying patterns, capturing non-linear relationships and rare events better than traditional methods.

In this paper, we employ a hybrid data augmentation strategy to address the limitations of relying solely on GAN-generated synthetic data in failure detection models. While GANs offer the advantage of generating diverse data, they are prone to mode collapse and distribution bias, where the synthetic data fails to capture the full spectrum of real-world failure modes. These issues can skew model predictions and reduce accuracy. To mitigate mode collapse, we carefully monitor the generator and discriminator loss to detect early signs of mode collapse. We also use noise regularization to introduce variation in the training process, which discourages the generator from converging to a narrow set of outputs.

Instead of fully relying on GAN-generated data, we created training datasets that combine both real and synthetic samples in carefully controlled proportions. This hybrid approach ensures that the essential characteristics of actual AFC system failures are preserved while leveraging the variability introduced by synthetic data. The combination reduces over-reliance on synthetic data, enhances the model's generalization capability, and prevents overfitting. The effectiveness of this approach is supported by Yuan et al., who demonstrated the benefits of combining real and synthetic data to improve model performance in failure detection scenarios [50].

The original dataset consisted of 518 fault records, which were reduced to 36 after data cleaning due to the removal of incomplete or inconsistent entries. Following data augmentation, the dataset increased to 1398 records. To ensure generalization, the model's performance was validated primarily using real-world data, with synthetic data only supporting the training phase. By regularly validating on real data, we ensured that the model maintains robustness in real-world applications and avoids overfitting to the characteristics of the synthetic samples.

The results in Table 3 clearly show that data augmentation methods can significantly influence model performance. Weibull-GAN was the most effective, considerably reducing both MAE and MSE, indicating its superiority in generating syn-thetic data for enhancing model accuracy. In contrast, SMOTE worsened the performance, while GAN offered moderate improvements over no augmentation.

Table 3. Model performance with different data augmentation techniques.

Model	MAE	MSE
No data augmentation	0.03952	0.00211
GAN	0.03618	0.00193
SMOTE	0.06617	0.00532
Weibull-GAN	0.02869	0.00111

To evaluate the impact of different GAN noise dimensions and Weibull distribution parameters on model performance, a sensitivity analysis was conducted. We varied the noise dimension of the GAN and the shape and scale parameters of the Weibull distribution, and assessed the corresponding changes in the MAE and MSE. As shown in Table 4, the combination of a noise dimension of 500, a Weibull shape parameter of 1.0, and a scale parameter of 500 yielded the best model performance, achieving the lowest MAE (0.02869) and MSE (0.00111). This indicates that with these specific parameters, the model was able to generate data that most closely resembled the real data, minimizing prediction error. In contrast, lower noise dimensions or different Weibull parameters resulted in higher error rates, demonstrating the model's sensitivity to these key hyperparameters. This

analysis highlights the importance of proper parameter selection to optimize the predictive capabilities of the model.

Table 4. Sensitivity analysis of performance with varying noise dimensions and Weibull distribution parameters.

GAN Noise Dimension	Weibull Shape	Weibull Scale	MAE	MSE
500	2.0	400	0.03513	0.00194
100	1.0	400	0.03753	0.00241
500	1.0	500	0.02869	0.00111

4.2. Health Assessment of AFC System on Line 10

In this paper, expert bias in weight assignment was minimized through the integration of the AHP with the critic entropy method. This hybrid approach combines subjective judgments from experts with objective data-driven methods, thus balancing expert input with measurable variability in criteria. AHP was used to establish a consistency matrix to prioritize factors based on expert judgment, while the critic entropy method was applied to calculate objective weights, reducing over-reliance on subjective opinions. By using entropy to quantify the variability and informativeness of each criterion, this approach enhances the transparency and objectivity of the decision-making process. Previous research supports this combined method. For example, Tan and Li showed that the consensus matrix generated by entropy helps mitigate expert bias, while Chen demonstrated that the integration of entropy and AHP improves decision-making accuracy by balancing subjective and objective weights [51,52].

4.2.1. Subjective Weight Judgment Matrix

In this paper, expert weight assignment was conducted using a questionnaire-based approach combined with the AHP. This allowed for structured expert opinion collection and the construction of comparative weight matrices. The AHP method ensures transparency and reduces bias by systematically comparing factors and assigning weights based on quantifiable expert judgments. To minimize potential expert bias in the AHP-CRITIC entropy-based health assessment, we consulted a diverse panel of twelve experts. This group included five senior engineers in urban rail transit with extensive experience in AFC systems, four academic researchers specializing in prognostics and health management, and three maintenance specialists from Shanghai Shentong Metro Group with practical expertise in AFC equipment. This balanced mix of academic and industrial perspectives provided objective and comprehensive input in the assessment, reducing the influence of subjective judgments.

The weight matrices constructed are system health degree judgement matrix A, fault site, spare parts usage, relative importance of site matrix B_1, Relative importance of spare parts matrix B_2, fault impact time matrix B_3, fault level matrix B_4, respectively:

$$A = \begin{bmatrix} 1 & \frac{1}{2} & 1 & \frac{1}{4} \\ 2 & 1 & 2 & \frac{1}{4} \\ 1 & \frac{1}{2} & 1 & \frac{1}{3} \\ 4 & 4 & 3 & 1 \end{bmatrix} \tag{21}$$

$$B_1 = \begin{bmatrix} 1 & 3 & 5 & 4 \\ \frac{1}{3} & 1 & 4 & 3 \\ \frac{1}{5} & \frac{1}{4} & 1 & 2 \\ \frac{1}{4} & \frac{1}{3} & \frac{1}{2} & 1 \end{bmatrix} \tag{22}$$

$$B_2 = \begin{bmatrix} 1 & \frac{1}{2} & 2 & 4 & 2 & 3 & 2 \\ 2 & 1 & 5 & 2 & 5 & 5 & 5 \\ \frac{1}{2} & \frac{1}{5} & 1 & 3 & \frac{1}{2} & 3 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{3} & 1 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{5} & 2 & 3 & 1 & 3 & 2 \\ \frac{1}{3} & \frac{1}{5} & \frac{1}{3} & 3 & \frac{1}{3} & 1 & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{5} & 2 & 3 & \frac{1}{2} & 3 & 1 \end{bmatrix} \tag{23}$$

$$B_3 = \begin{bmatrix} 1 & \frac{1}{8} \\ 8 & 1 \end{bmatrix} \tag{24}$$

$$B_4 = \begin{bmatrix} 1 & 3 & 5 & 9 \\ \frac{1}{3} & 1 & 2 & 3 \\ \frac{1}{5} & \frac{1}{2} & 1 & 4 \\ \frac{1}{9} & \frac{1}{3} & \frac{1}{4} & 1 \end{bmatrix} \tag{25}$$

The consistency test of the five matrices was carried out, and the obtained consistency ratios were: 0.0363, 0.0787, 0.0977, 0.0000, 0.0395 were all less than 0.1, which met the consistency requirement, indicating that the four judgement matrices were acceptable.

4.2.2. Calculation of Subjective and Objective Fusion Weights

The weights of each factor were obtained by calculating the eigenvectors of the weight comparison matrix, which were subsequently normalized to ensure that all weights summed to 1. The objective weights are shown as ω_i in Table 5.

Table 5. Characteristic objective weights.

Indicator	Weight	Indicator	Weight	Indicator	Weight
Fault level _Primary Fault	0.3163	Spares importance _Battery	0.0387	Fault impact time _Less than 30 min	0.0148
Fault impact time _More than 30 min	0.1184	Part importance _Host	0.0331	Part importance _Network device	0.0141
Fault level _Secondary Fault	0.1108	Fault level _Other fault	0.0302	Spares importance _Motherboard	0.0132
Fault level _Tertiary fault	0.0778	Spares importance _NIC	0.0279	Spares importance _Memory	0.0117
Spares importance _Hard disk	0.0745	Spares importance _Controller	0.0237	Part importance _Encoder	0.0110
Part importance _Storage	0.0637	Spares importance _Capacitor	0.0200		

Subsequently, subjective ω_j weights were calculated using the critic entropy weighting method, and the subjective weights obtained were shown in Table 6.

Table 6. Characteristic subjective weights.

Indicator	Weight
Component fault rate	0.197140409
Fault grade	0.268415689
Spare parts utilization rate	0.265786498
Fault impact time	0.268657404

Based on these weights, the health level of the AFC system is shown in Figure 7:

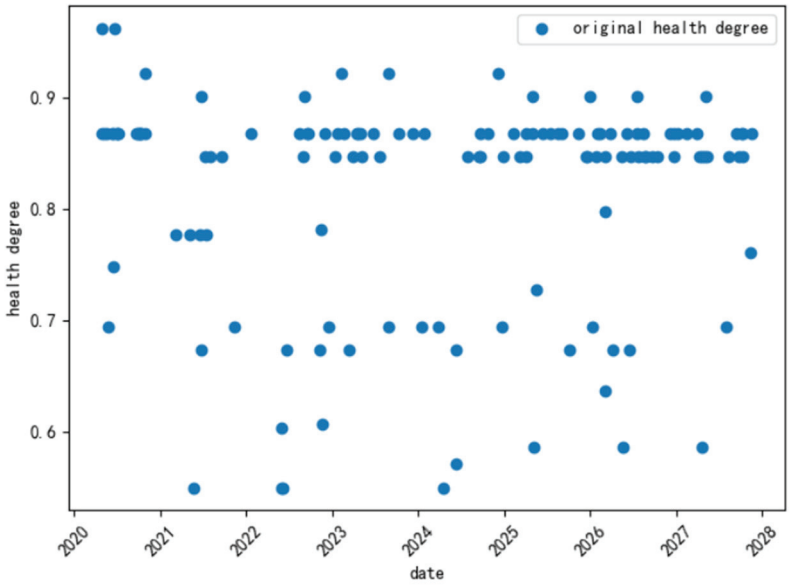


Figure 7. Health degree of AFC system based on weight calculation.

The health level subsequently calculated in fusion with the ADM model is shown in Figure 8:

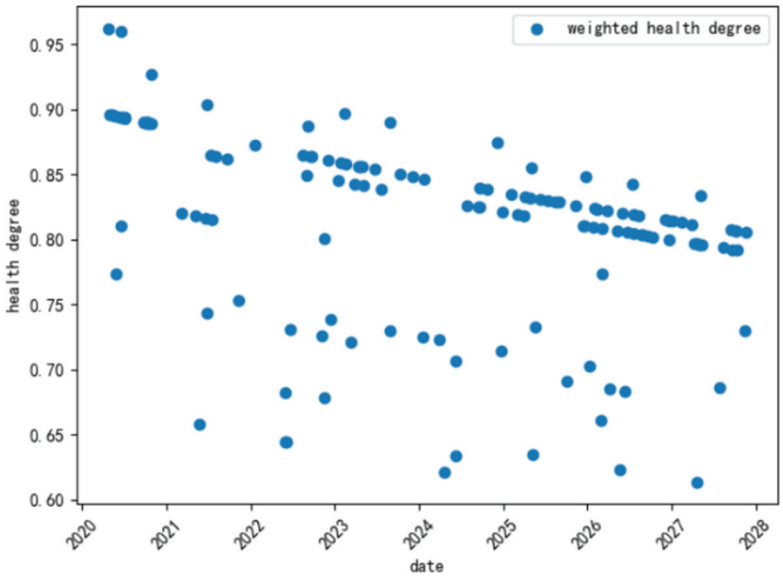


Figure 8. Health degree of AFC system based on digital analog fusion.

4.3. Prediction of Remaining Life of AFC System on Line 10

Figures 9 and 10 reflect the model prediction of the future health value and health level changes of the AFC system.

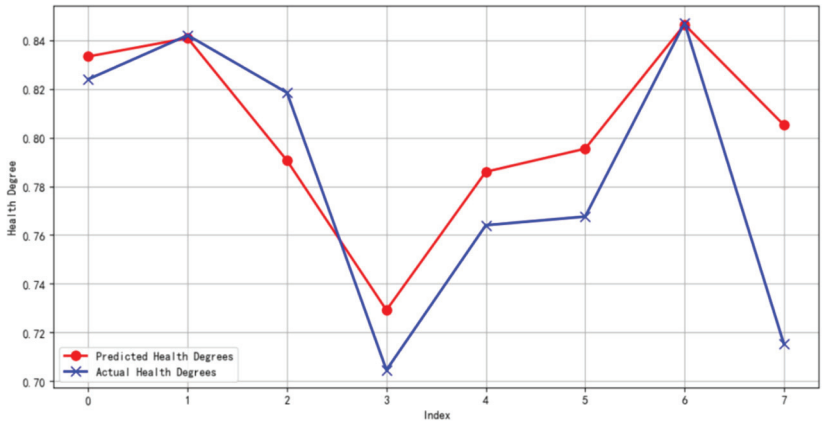


Figure 9. Comparison of predicted health values with actual health values (validation set).

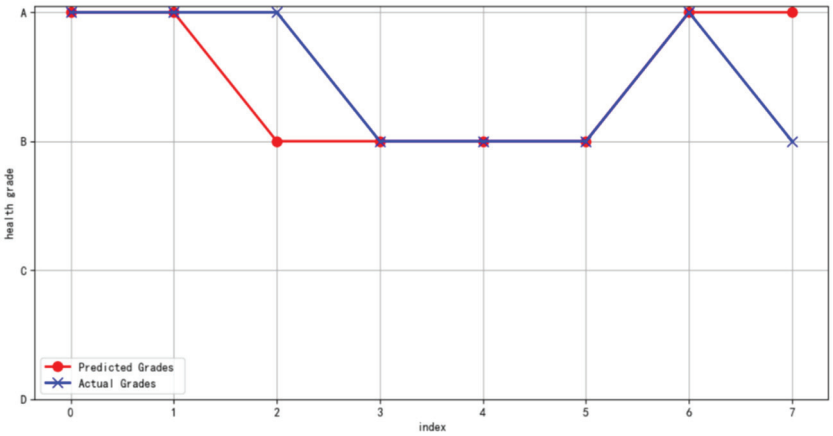


Figure 10. Comparison of prediction health grade and actual health grade (validation set).

To show the performance of SSA-XGBoost, this paper compared it with other models; the results are shown in Table 7. Figures 11–13 demonstrate the prediction performance of each model more intuitively. By comparing the performance of these models on the same dataset, the SSA-XGBoost model shows excellent performance in the key performance indicators such as MSE, MAE and R-squared. Especially on the validation set, the MSE and MAE values of the SSA-XGBoost model are much lower than other models, showing its advancement in processing complex data and pattern recognition. The SSA configuration includes a population size of 100, representing the number of candidate solutions initialized in the search space. A larger population size increases search diversity, though at the cost of higher computational complexity. The maximum number of iterations is set to 200. To address operational variances, the SSA is used to optimize key XGBoost hyperparameters, including max_depth, n_estimators, and learning_rate, by tuning them specifically for different maintenance intervals in metro systems. For systems with shorter maintenance intervals, a relatively high learning_rate of around 0.1–0.3 helps the model quickly adapt to rapid changes in the system. Conversely, for metro systems with longer maintenance intervals, increasing n_estimators to values around 300–500 and setting max_depth to a range of 6–10 allows the model to better capture slower-developing, gradual faults. These specific parameter values improve the model's adaptability, enabling it to perform effectively across various operational rhythms in different metro systems.

Table 7. Evaluation results of different models.

Model	Test Set MSE	Test Set MAE	Validation Set MSE	Validation Set MAE
XGBoost	0.00178	0.03006	0.00359	0.04965
Random Forest	0.00158	0.03872	0.00413	0.05238
Linear Regression	0.51118	0.52810	0.01371	0.10514
SSA-XGBoost	0.00111	0.02869	0.00103	0.02879

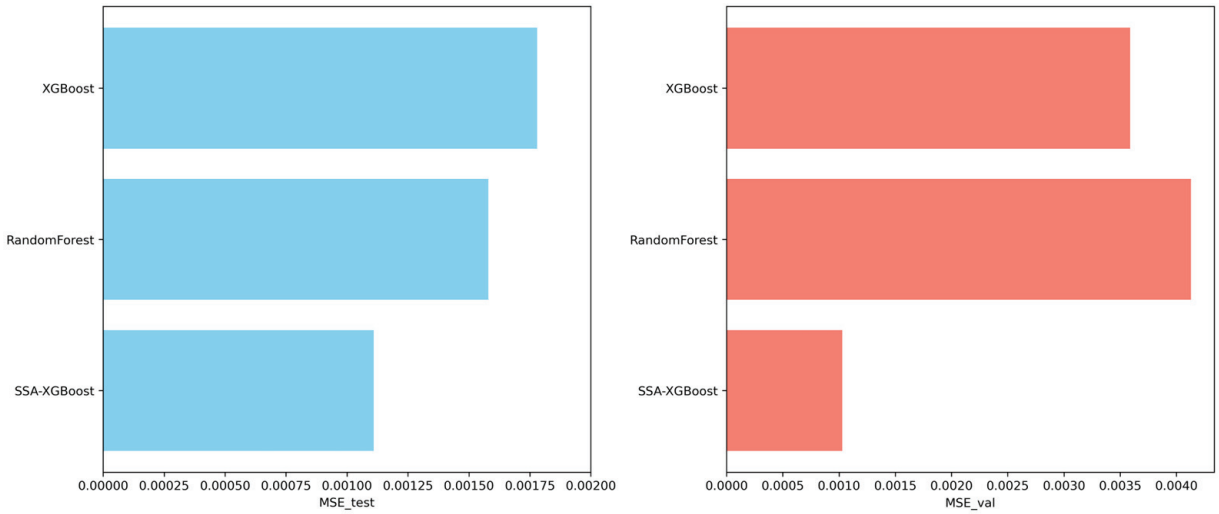


Figure 11. Mean square error for different models.

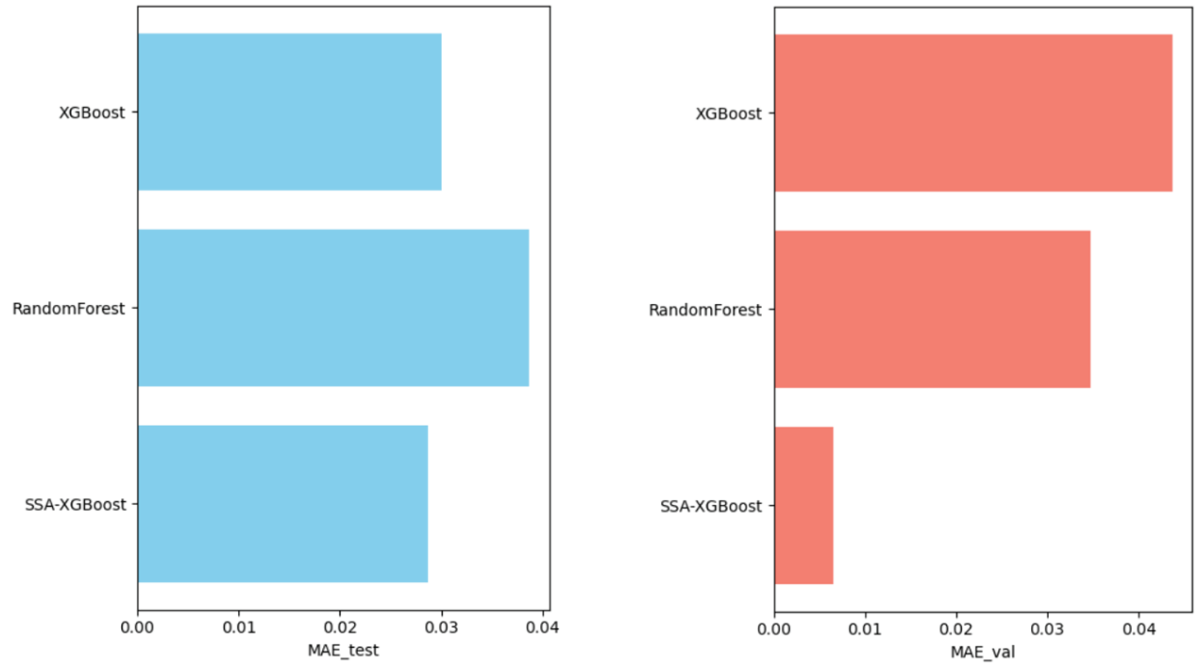


Figure 12. Mean absolute errors for different models.

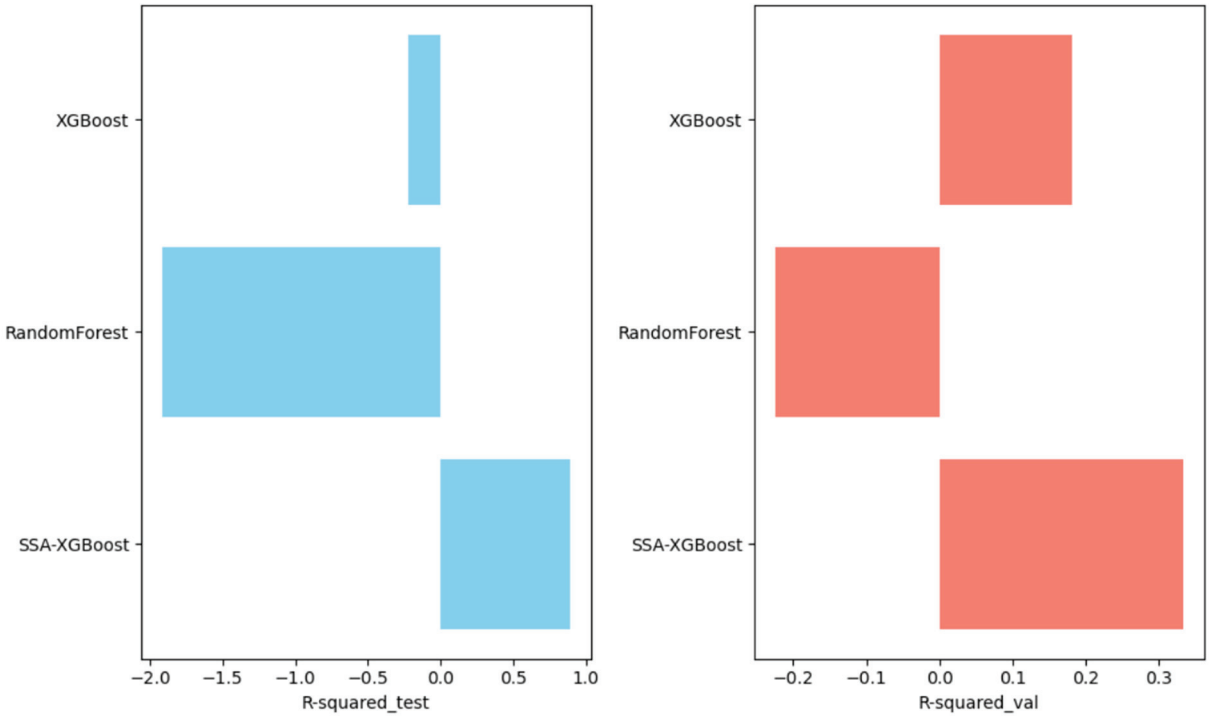


Figure 13. R-squared for different models.

We utilized permutation feature importance (PFI) to assess the significance of various features in our predictive model. We utilized the mean squared error (MSE) loss values to evaluate the significance of various features in our predictive model. The results of the feature importance analysis are illustrated in Figure 14.

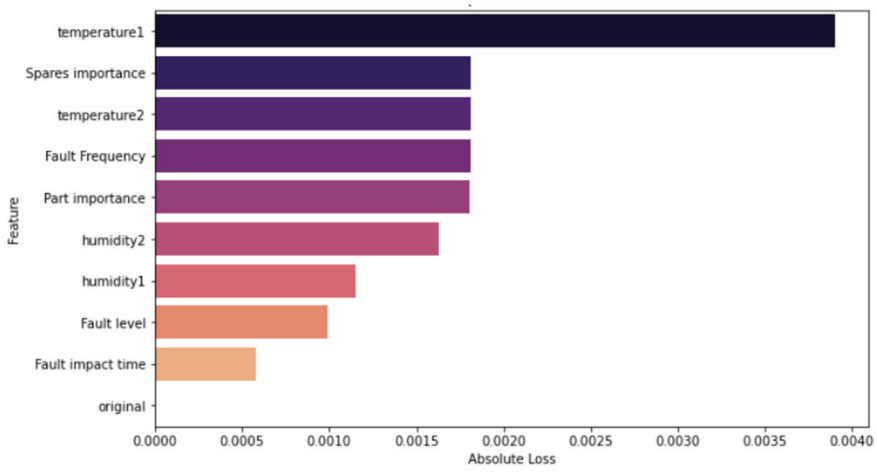


Figure 14. Feature importance based on PFI.

The results of this analysis reveal that different features contribute unevenly to the model's performance, highlighting the importance of certain variables over others. Specifically, features such as *temperature* and *humidity* exhibit the highest importance scores, indicating their substantial influence on the model's predictive accuracy. This suggests that fluctuations in these environmental parameters play a critical role in determining the outcomes of our model.

The ablation experiments conducted in this paper are shown in Table 8. The results clearly demonstrate the effectiveness of the research method.

Table 8. Results of ablation experiments.

Model	MSE	MAE	R2_Score
No feature extraction XGBoost	0.01599	0.11832	−5.77417
LSTM feature extraction + XGBoost	0.00178	0.03006	−2.91019
SSA-XGBoost without feature extraction	0.00163	0.03460	0.33736
LSTM feature extraction + SSA-XGBoost	0.00111	0.02869	0.50421

The SSA-XGBoost model demonstrates notable advantages for predicting RUL in AFC systems due to its capacity to handle complex, sparse, and multi-dimensional data relationships. By incorporating multi-layer LSTM networks for feature extraction, SSA-XGBoost captures long-term dependencies and nonlinear patterns present in AFC data, such as fault data and environmental factors, which are often overlooked by simpler models. Additionally, the Sparrow Search Algorithm (SSA) optimizes the XGBoost model through adaptive global and local searches, preventing local optima and reducing overfitting, which are critical for the irregular and sparse characteristics of AFC data. Furthermore, advanced data augmentation techniques using GAN and Weibull distributions enhance the dataset's diversity and relevance, making the model more resilient to weak and sparse features. Finally, SSA-XGBoost's adaptive optimization enables the model to dynamically adjust to changing data patterns, maintaining high predictive accuracy across varying conditions. These features make SSA-XGBoost particularly well-suited for the AFC system, surpassing simpler models like Random Forest that lack the same level of dynamic optimization and data-handling flexibility.

5. Conclusions

This paper addresses the health assessment and prediction of remaining useful life of automated fare collection (AFC) system, proposes a digital analog fusion-based health status assessment method for AFC, and establishes the SSA-XGBoost prediction model to achieve the prediction of RUL of AFC electromechanical equipment under sparse data. The results of this paper show the following:

(1) Compared with a single data-driven or model-driven method, the health assessment method proposed in this paper can determine the health level of the equipment in terms of the assessment results, thus guiding the development of differentiated operation and maintenance strategies, and providing refined and intelligent maintenance and repair solutions for equipment maintenance personnel.

(2) Multi-layer LSTM with multi-feature extraction, as through deep learning and the extraction of multi-dimensional features such as temperature and humidity, fault types, parts, etc., allow the model to capture long-term time-dependent and multi-dimensional feature interactions, solving the difficulty of sparse raw data and weak features and making prediction results more accurate and reliable.

(3) The XGBoost model optimized with SSA shows excellent performance in several performance indicators. Especially, MSE of the model is 0.00111 and MAE is 0.02869 in the test set, while the MSE is reduced to 0.00004 and the MAE is reduced to 0.00659 in the validation set, which are significantly better than the other comparative models, such as XGBoost, random forest regression and linear regression. In addition, the SSA-XGBoost model also performs well on R-squared, which further verifies its effectiveness in prediction accuracy and model fitting. The SSA algorithm provides global and local search capabilities for XGBoost, and it can improve the ability of the XGBoost model in RUL prediction.

Ref. [53] provides a comparative examination of the proposed model for evaluating the health of AFC system against the methodology developed by Qi Doudou, which emphasizes the prediction of RUL. The proposed model offers several notable advantages, particularly in its incorporation of sophisticated machine learning techniques, such as a multi-layer LSTM architecture paired with SSA-XGBoost. This combination significantly improves predictive accuracy by effectively capturing complex temporal dependencies within the data. Additionally, the model integrates a wide range of operational and environmental data sources, which contributes to a more comprehensive assessment of system health.

The model developed in this paper, although applied specifically to the AFC system of Shanghai metro line 10, demonstrates significant potential for adaptation to other metro and railway systems. The core components, including the multi-layer LSTM architecture for feature extraction and the SSA-XGBoost for remaining useful life prediction, are designed to be data-driven and flexible, allowing for seamless application to diverse systems with adequate operational data. The modular framework enables easy integration of various data types, while the dynamic parameter optimization provided by the SSA-XGBoost ensures that the model can be fine-tuned to meet the operational characteristics of different systems. However, in cases where AFC systems exhibit unique configurations or distinct fault behaviors, adjustments to the data preprocessing pipeline and model retraining may be necessary to ensure optimal performance. Overall, this research provides a scalable and generalizable solution for AFC system health assessment and RUL prediction, with the capability to be effectively extended to other operational environments through system-specific modifications where required. Although the SSA-XGBoost model has demonstrated high accuracy and adaptability in this study, we have not yet conducted validation tests in other metro systems. Consequently, while the model shows promise for broader application, its deployment across diverse metro systems may encounter several potential challenges. Firstly, data quality and availability can vary significantly among metro systems; some systems may lack complete historical fault data or environmental monitoring records, which could impact feature extraction and model performance. Additionally, differences in AFC equipment types and maintenance practices could lead to variations in fault patterns and maintenance needs, potentially necessitating retraining or fine-tuning of key model components, such as the LSTM feature extraction layers. Finally, variations in environmental factors and passenger volume could impact equipment health differently, with systems operating in diverse climates or under varying usage levels introducing new data trends or outliers. Therefore, for SSA-XGBoost to be effectively applied in other metro systems, targeted modifications to data preprocessing and model adaptability would be essential, along with preliminary testing to ensure reliability and accuracy in new contexts.

In future research, we plan to conduct pilot tests in other metro systems to validate and refine the model's adaptability and performance across different operational environments. Through the demonstration application in Shanghai Metro Line 10, this paper can effectively deal with the challenges of complexity and diversity of AFC systems, data sparsity and weak features, and demonstrates higher accuracy and reliability than traditional health assessment and RUL prediction methods. In the future, from the data level of fusion sensor monitoring variable information, in the extraction of eigenvalues, can also be other aspects of the eigenvalue expansion, such as time domain, time-frequency domain entropy features, etc., to extract more eigenvalues that may contain the performance of the AFC equipment state information. Because of the huge number of AFC equipment components, the RUL of the system is affected by different components and some of the components also have the characteristics of gradual fault type. In the future, we can carry out degradation experiments of key components, and at the same time, we can consider the nonlinear

problems in the process of AFC equipment state change to improve the performance and universality of the method. The failure of AFC equipment is also directly related to human behavior. In the future, we will collect data related to ergonomic risk factors and expand the consideration of ergonomic risk factors. We also consider the incorporation of advanced observer-based control strategies, such as the Takagi–Sugeno fuzzy model, to optimize model and design the feature importance assesses method.

Author Contributions: Conceptualization, J.X. and G.Y.; methodology, G.Y. and J.X.; formal analysis, Y.S. and J.X.; investigation, Z.X. and Y.W.; resources, Y.W.; data curation, Y.W. and Z.X.; verification, Z.X. and J.X.; writing—original draft preparation, Z.X., G.Y. and J.X.; writing—review and editing, Y.S., G.Y. and Z.X.; visualization, G.Y. and Z.X.; supervision, J.X.; project administration, Y.W. and J.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Ministry of Industry and Information Technology of China (No. 2022-A04R-1-1) and the Chinese National Nature Science Foundation Project (No. 72471133) and Natural Science Foundation of Shanghai Municipality (No. 21ZR1423800).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are confidential and were obtained from Shanghai Shentong Metro Group Co., Ltd.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Guofu, W.; Hongchen, W.; Haidong, L. Application and Prospect for New Technology of AFC system in Urban Rail Transit. *Urban Rapid Rail Transit* **2017**, *30*, 41–44.
2. Hora, J.; Dias, T.G.; Camanho, A.S.; Sobral, T. Estimation of Origin-Destination Matrices Under Automatic Fare Collection: The Case Study of Porto Transportation System. *Transp. Res. Procedia* **2017**, *27*, 664–671. [CrossRef]
3. Foote, P.J.; Patronsky, R.T.; Stuart, D.G. Transit Customer Acceptance of Automated Fare Collection Systems. *J. Public Transp.* **1999**, *2*, 1–20. [CrossRef]
4. Bieler, M.; Skretting, A.; Büdinger, P.; Grønli, T.-M. Survey of Automated Fare Collection Solutions in Public Transportation. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 14248–14266. [CrossRef]
5. Liu, Z.; Liu, S. Operation support of Beijing metro automatic fare collection system clearance house system. In Proceedings of the 2016 International Conference on Logistics, Informatics and Service Sciences (LISS), Sydney, NSW, Australia, 24–27 July 2016.
6. Liu, Z.L.; Liu, S.F.; Chen, T.D. In Metro Automatic Fare Collection System Safety and Security. In Proceedings of the 19th International Conference on Electromagnetics in Advanced Applications (ICEAA), Verona, Italy, 11–15 September 2017; IEEE: Verona, Italy, 2017; pp. 604–607.
7. Hu, Y.; Miao, X.W.; Si, Y.; Pan, E.S.; Zio, E. Prognostics and health management: A review from the perspectives of design, development and decision. *Reliab. Eng. Syst. Saf.* **2022**, *217*, 15. [CrossRef]
8. Yin, X.X.; He, W.; Cao, Y.; Ma, N.; Zhou, G.H.; Li, H.Y. A new health state assessment method based on interpretable belief rule base with bimetric balance. *Reliab. Eng. Syst. Saf.* **2024**, *242*, 17. [CrossRef]
9. Zhou, Z.J.; Feng, Z.C.; Hu, C.H.; Hu, G.Y.; He, W.; Han, X.X. Aeronautical relay health state assessment model based on belief rule base with attribute reliability. *Knowl.-Based Syst.* **2020**, *197*, 13. [CrossRef]
10. Gucik, D.; Zolghadri, A.; Suhir, E.; Bechou, L. A model-based prognosis strategy for prediction of Remaining Useful Life of Ball-Grid-Array Interconnections. *IFAC Proc. Vol. (IFAC-PapersOnline)* **2014**, *19*, 7354–7360. [CrossRef]
11. Tang, Y.; Yang, X.; Wang, G.R. A Framework for Health State Evaluation of the Complex Mechanical System With Its Occurrence Probability of Failure Mode. *IEEE Access* **2020**, *8*, 73570–73587. [CrossRef]
12. Zhang, Y.; Li, Y.-F.; Zhang, M.; Wang, H. A novel health indicator by dominant invariant subspace on Grassmann manifold for state of health assessment of lithium-ion battery. *Eng. Appl. Artif. Intell.* **2024**, *130*, 107698. [CrossRef]
13. Zhang, Y.; Sun, J.H.; Zhang, J.; Shen, H.R.; She, Y.Y.; Chang, Y. Health state assessment of bearing with feature enhancement and prediction error compensation strategy. *Mech. Syst. Signal Proc.* **2023**, *182*, 17. [CrossRef]
14. Guo, Q.; Qui, W.; Deng, H.R.; Zhang, X.Y.; Li, Y.; Wang, X.W.; Yan, X.W. In Estimation of Electric Vehicle Battery State of Health based on Relative State of Health Evaluation. In Proceedings of the 2nd IEEE Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, 25–26 March 2017; IEEE: Chongqing, China, 2017; pp. 1998–2002.

15. Lee, J.H.; Kim, H.S.; Lee, I.S. In State of Charge Estimation and State of Health Diagnostic Method Using Multilayer Neural Networks. In Proceedings of the 20th International Conference on Electronics, Information, and Communication (ICEIC), Jeju, Republic of Korea, 31 January–3 February 2021; IEEE: Jeju, Republic of Korea, 2021.
16. Zhang, L.L.; Qiao, F.; Wang, J.K.; Zhai, X.D. Equipment Health Assessment Based on Improved Incremental Support Vector Data Description. *IEEE Trans. Syst. Man Cybern.-Syst.* **2021**, *51*, 3205–3216. [CrossRef]
17. Zhou, Y.B.; Xiao, H.; Chen, Z.G.; Wang, T.; Pan, Y.H.; Wan, Y.B. A Dynamic Health Assessment Method for Industrial Equipment Based on SG-FCM Clustering Algorithm. In Proceedings of the 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Atlanta, GA, USA, 14–17 July 2019; IEEE: Atlanta, GA, USA, 2019; pp. 34–39.
18. Jiang, Z.; Fang, H.; Wang, X.; Fan, H. Health Evaluation Method of Satellite Battery Based-On Bayesian Network. In Proceedings of the 2019 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC), Beijing, China, 15–17 August 2019; pp. 552–556.
19. Mi, J.H.; Li, Y.F.; Yang, Y.J.; Peng, W.W.; Huang, H.Z. Reliability assessment of complex electromechanical systems under epistemic uncertainty. *Reliab. Eng. Syst. Saf.* **2016**, *152*, 1–15. [CrossRef]
20. Son, J.B.; Zhang, Y.L.; Sankavaram, C.; Zhou, S.Y. RUL Prediction for Individual Units Based on Condition Monitoring Signals With a Change Point. *IEEE Trans. Reliab.* **2015**, *64*, 182–196. [CrossRef]
21. Lei, Y.G.; Li, N.P.; Gontarz, S.; Lin, J.; Radkowski, S.; Dybala, J. A Model-Based Method for Remaining Useful Life Prediction of Machinery. *IEEE Trans. Reliab.* **2016**, *65*, 1314–1326. [CrossRef]
22. Zhang, C.Y.; Wang, C.S.; Lu, N.Y.; Jiang, B. An RBMs-BN method to RUL prediction of traction converter of CRH2 trains. *Eng. Appl. Artif. Intell.* **2019**, *85*, 46–56. [CrossRef]
23. Ardeshiri, R.R.; Ma, C.B. Multivariate gated recurrent unit for battery remaining useful life prediction: A deep learning approach. *Int. J. Energy Res.* **2021**, *45*, 16633–16648. [CrossRef]
24. Zhang, Q.; Ye, Z.J.; Shao, S.Y.; Niu, T.L.; Zhao, Y.W. Remaining useful life prediction of rolling bearings based on convolutional recurrent attention network. *Assem. Autom.* **2022**, *42*, 372–387. [CrossRef]
25. Ren, L.; Sun, Y.Q.; Cui, J.; Zhang, L. Bearing remaining useful life prediction based on deep autoencoder and deep neural networks. *J. Manuf. Syst.* **2018**, *48*, 71–77. [CrossRef]
26. Zhang, Z.Z.; Song, W.; Li, Q.Q. Dual-Aspect Self-Attention Based on Transformer for Remaining Useful Life Prediction. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 11. [CrossRef]
27. Chang, Z.H.; Yuan, W.; Huang, K. Remaining useful life prediction for rolling bearings using multi-layer grid search and LSTM. *Comput. Electr. Eng.* **2022**, *101*, 13. [CrossRef]
28. An, Q.L.; Tao, Z.R.; Xu, X.W.; El Mansori, M.; Chen, M. A data-driven model for milling tool remaining useful life prediction with convolutional and stacked LSTM network. *Measurement* **2020**, *154*, 15. [CrossRef]
29. Wang, M.X.; Xiang, G.; Cui, L.F.; Zhang, Q.Z.; Chen, J. Remaining useful life distribution prediction framework for lithium-ion battery fused prior knowledge and monitoring data. *Meas. Sci. Technol.* **2023**, *34*, 125108. [CrossRef]
30. Chen, W.B.; Liu, C.; Chen, W.Z.; Liu, H.X.; Chen, Q.L.; Wu, P.L. A Prediction Method for the RUL of Equipment for Missing Data. *Complexity* **2021**, *2021*, 10.
31. Mo, H.; Iacca, G. Evolutionary neural architecture search on transformers for RUL prediction. *Mater. Manuf. Process.* **2023**, *38*, 1881–1898. [CrossRef]
32. Youness, G.; Aalah, A. An Explainable Artificial Intelligence Approach for Remaining Useful Life Prediction. *Aerospace* **2023**, *10*, 474. [CrossRef]
33. Chao, M.A.; Kulkarni, C.; Goebel, K.; Fink, O. Fusing physics-based and deep learning models for prognostics. *Reliab. Eng. Syst. Saf.* **2022**, *217*, 18.
34. Deng, Y.; Du, S.; Wang, D.; Shao, Y.; Huang, D. A Calibration-Based Hybrid Transfer Learning Framework for RUL Prediction of Rolling Bearing Across Different Machines. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 3511015. [CrossRef]
35. Wahid, A.; Yahya, M.; Breslin, J.G.; Intizar, M.A. Self-Attention Transformer-Based Architecture for Remaining Useful Life Estimation of Complex Machines. *Procedia Comput. Sci.* **2023**, *217*, 456–464. [CrossRef]
36. Wang, F.K.; Mamo, T. Hybrid approach for remaining useful life prediction of ball bearings. *Qual. Reliab. Eng. Int.* **2019**, *35*, 2494–2505. [CrossRef]
37. Liao, L.X.; Köttig, F. A hybrid framework combining data-driven and model-based methods for system remaining useful life prediction. *Appl. Soft. Comput.* **2016**, *44*, 191–199. [CrossRef]
38. Hu, X.X.; Tan, L.; Tao, T. M2BIST-SPNet: RUL prediction for railway signaling electromechanical devices. *J. Supercomput.* **2024**, *80*, 1–31. [CrossRef]
39. Mudiyanselage, S.E.; Nguyen, P.H.D.; Rajabi, M.S.; Akhavian, R. Automated Workers' Ergonomic Risk Assessment in Manual Material Handling using sEMG Wearable Sensors and Machine Learning. *arXiv* **2021**, arXiv:2109.15036. [CrossRef]

40. Halimi, Z.; Bavafa, A.; Cui, Q. Barriers to Community Connectivity: An Assessment of Reconnecting Communities Pilot Program. *Int. Conf. Transp. Dev.* **2024**, *2024*, 83–91.
41. Ganesh, K.; Shuib, B.; Imam, A.A.; Khowaja, S.A.; Capretz, L.F.; Balogun, A.O. Data Harmonization for Heterogeneous Datasets: A Systematic Literature Review. *Appl. Sci.* **2021**, *11*, 8275. [CrossRef]
42. Zhang, D.D.; Yin, C.C.; Zeng, J.C.; Yuan, X.; Zhang, P. Combining structured and unstructured data for predictive models: A deep learning approach. *BMC Med. Inf. Decis. Mak.* **2020**, *20*, 280. [CrossRef] [PubMed]
43. Akshay, S.; William, W.; Ujwal, S.; Edward, T.; Bhargava-Shah, A.; Wu, J.Y.; Ren, A.L.; Caoili, K.; Bui, B.; Alkhani, L.; et al. Selective prediction for extracting unstructured clinical data. *J. Am. Med. Inform. Assoc.* **2024**, *31*, 188–197.
44. Liu, K.B.; Huang, S. Integration of Data Fusion Methodology and Degradation Modeling Process to Improve Prognostics. *IEEE Trans. Autom. Sci. Eng.* **2016**, *13*, 1–11. [CrossRef]
45. Zhang, Z.X.; Si, X.S.; Hu, C.H.; Lei, Y.G. Degradation data analysis and remaining useful life estimation: A review on Wiener-process-based methods. *Eur. J. Oper. Res.* **2018**, *271*, 775–796. [CrossRef]
46. Yang, Y.; Hu, X.; Qing, D.; Chen, F. Arrhenius Equation-Based Cell-Health Assessment: Application to Thermal Energy Management Design of a HEV NiMH Battery Pack. *Energies* **2013**, *6*, 2709–2725. [CrossRef]
47. Saman, T.; Morteza, A.; Sadeghi, S. Deep Learning Hyperparameter Optimization: Application to Electricity and Heat Demand Prediction for Buildings. *Energy Build.* **2023**, *289*, 113036.
48. Halimi, Z.; SafariTaherkhani, M.; Cui, Q.A. A Generalized Framework for Assessing Equity in Ground Transportation Infrastructure: An Exploratory Study. *arXiv* **2024**, arXiv:2409.19018.49. [CrossRef]
49. Yang, C.Y.; Shen, Y.J.; Xu, Y.H.; Zhao, D.L.; Dai, B.; Zhou, B.L. Improving GANs with A Dynamic Discriminator. *arXiv* **2022**, arXiv:2209.09897.
50. Yuan, F.F.; Shang, Y.; Liu, Y.B.; Tan, J.L. Data Augmentation for Insider Threat Detection with GAN. In Proceedings of the 2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI), Baltimore, MD, USA, 9–11 November 2020; pp. 32–638.
51. Tan, Z.W.; Li, P.Y. Group Decision-Making Information Security Risk Assessment Based on AHP and Information Entropy. *Res. J. Appl. Sci. Eng. Technol.* **2012**, *7*, 1–8.
52. Chen, C.-H. A Novel Multi-Criteria Decision-Making Model for Building Material Supplier Selection Based on Entropy-AHP Weighted TOPSIS. *Entropy* **2020**, *22*, 259. [CrossRef]
53. Qi, D.D. Exploration of Health Assessment Scheme for AFC Terminal Equipment Based on Remaining Life Prediction. *Technol. Innov. Appl.* **2022**, *4*, 135–138.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

smart cities

MDPI

Article

Real-Time Monitoring of Environmental Parameters in Schools to Improve Indoor Resilience Under Extreme Events

Salit Azoulay Kochavi [1], Oz Kira [1,2,*] and Erez Gal [1,2]

[1] The Department of Civil and Environmental Engineering, Faculty of Engineering Sciences, Ben-Gurion University of the Negev, Beer-Sheva 8410501, Israel; salitkochavi@gmail.com (S.A.K.); erezgal@bgu.ac.il (E.G.)
[2] The Goldman Sonnenfeldt School of Sustainability and Climate Change, Ben-Gurion University of the Negev, Beer-Sheva 8410501, Israel
* Correspondence: ozkira@bgu.ac.il

Highlights:

What are the main findings?

- The study emphasizes the critical need for real-time IAQ monitoring in educational facilities, especially during extreme weather events. By closely tracking air pollutants, particularly PM2.5, temperature, and humidity, the study reveals that indoor air quality can deteriorate significantly during such events, potentially impacting vulnerable children's health.
- The study observed that during sandstorms, indoor PM2.5 levels rose by over 16%, and temperatures increased by more than 5% compared to normal conditions. These findings underline the significant infiltration of outdoor pollutants indoors, even with windows and doors closed, and call for enhanced ventilation and filtration systems in schools.

What is the implication of the main finding?

- This research aligns with the objectives of smart cities by calling for intelligent, real-time IAQ monitoring in schools, especially those in regions susceptible to climate extremes. The study supports policy initiatives focused on implementing centralized ventilation, monitoring systems, and air quality regulations in school environments, enhancing urban resilience to environmental stressors.
- To mitigate health risks, the study suggests that policy frameworks establish indoor air quality guidelines for educational settings, especially in arid or polluted regions. It also highlights the importance of further research to refine IAQ models and recommends installing low-cost, routinely calibrated sensors for ongoing IAQ assessment, aiming to create safer, healthier indoor environments for children.

Abstract: Climatic changes lead to many extreme weather events throughout the globe. These extreme weather events influence our behavior, exposing us to different environmental conditions, such as poor indoor quality. Poor indoor air quality (IAQ) poses a significant concern in the modern era, as people spend up to 90% of their time indoors. Ventilation influences key IAQ elements such as temperature, relative humidity, and particulate matter (PM). Children, considered a vulnerable group, spend approximately 30% of their time in educational settings, often housed in old structures with poorly maintained ventilation systems. Extreme weather events lead young students to stay indoors, usually behind closed doors and windows, which may lead to exposure to elevated levels of air pollutants. In our research, we aim to demonstrate how real-time monitoring of air pollutants and other environmental parameters under extreme weather is important for regulating the indoor environment. A study was conducted in a school building with limited ventilation located

in an arid region near the Red Sea, which frequently suffers from high PM concentrations. In this study, we tracked the indoor environmental conditions and air quality during the entire month of May 2022, including an extreme outdoor weather event of sandstorms. During this month, we continuously monitored four classrooms in an elementary school built in 1967 in Eilat. Our findings indicate that PM2.5 was higher indoors (statistically significant) by more than 16% during the extreme event. Temperature was also elevated indoors (statistically significant) by more than 5%. The parameters' deviation highlights the need for better indoor weather control and ventilation systems, as well as ongoing monitoring in schools to maintain healthy indoor air quality. This also warrants us as we are approaching an era of climatic instability, including higher occurrence of similar extreme events, which urge us to develop real-time responses in urban areas.

Keywords: urban sustainability; air pollution; indoor air quality; climate change; extreme events

1. Introduction

As we enter an era of climatic instability, indoor air quality and environmental comfort face a heightened risk of deterioration. With extreme events becoming more frequent [1], the resilience of the indoor environment in maintaining its atmosphere becomes increasingly crucial [2]. This study focuses on the Middle East, a region historically characterized by extreme conditions, such as high temperatures, low precipitation, and dust storms, impacting human health. Recent climatic changes and ongoing desertification have led to high uncertainty and may result in more frequent and severe dust storms. These events are expected to become even more common as we move into an increasingly unstable climatic future.

In recent years, scientific evidence has highlighted that indoor air quality (IAQ) can be compromised [3]. Poor air quality can stem from various indoor air pollutants, including biological, chemical, and physical sources. It can adversely affect the occupants' health, leading to symptoms such as sneezing, teary eyes, coughing, shortness of breath, dizziness, fatigue, and digestive problems. Furthermore, it can exacerbate conditions such as asthma and angina and even lead to life-threatening illnesses or sudden death. Poor IAQ can also promote the spread of biological and chemical factors, impacting occupants' perception of stuffiness and comfort [4]. Given that people now spend up to 90% of their time indoors [5], the health risks associated with poor IAQ are becoming increasingly significant. Particularly vulnerable groups include children, the elderly, and individuals with pre-existing lung or cardiovascular conditions [6,7]. The quality of the indoor environment is influenced by various factors, including site conditions, climate, building design, mechanical systems, construction methods, sources of pollution, occupants, and external transportation [8,9].

Children spend nearly 30% of their day at school, with 70% of that time spent in classrooms. However, overcrowded classrooms and aging school buildings with inadequate planning and maintenance often lead to poor ventilation and insulation. This exposes this vulnerable population to increased risks [10–12], which may impact their health and reduce their academic performance [13]. Moreover, children's underdeveloped respiratory systems, breathing patterns, and close physical interactions put them at risk of potential long- and short-term health issues [14–17].

Understanding the relationship between school building characteristics, their surroundings, and indoor air pollutants is crucial due to its potential implications for health and well-being, especially under harsh conditions and extreme events. This underscores

the importance of investigating air quality in classrooms, as poor IAQ can lead to decreased attendance, hindered learning achievements, and reduced teacher productivity [8,16], and ultimately lead to health issues. Air quality concerns in school buildings are widespread globally. For instance, half of U.S. schools reported internal air quality issues, while in Finland, 75–85% of school occupants experienced comfort issues affecting their performance levels [9]. Proper monitoring and continuous management of IAQ, such as by ensuring proper ventilation, can prevent or mitigate these issues. Additionally, adequate ventilation can reduce the spread of infectious diseases and enhance students' and teachers' health and performance [8,18,19].

To estimate the influence of extreme events on indoor environmental conditions, we tracked four measures in an elementary school: temperature (T), relative humidity (RH), carbon dioxide (CO_2), and fine particulate matter (PM2.5). These measures are influenced by external environmental conditions and the type and effectiveness of ventilation in the building [8,20]. Maintaining stable thermal conditions by avoiding frequent fluctuations in temperature and relative humidity, ensuring proper air mixing through ventilation systems, and adequate insulation are crucial for the thermal comfort of young children [7,21]. Poor maintenance of buildings and equipment can significantly contribute to poor indoor air quality, leading to discomfort and adverse health effects among occupants [17].

A relative humidity (RH) level greater than 60% can lead to moisture damage and promote the growth of microorganisms, exacerbating respiratory issues such as asthma symptoms in students and teachers [22]. Conversely, an RH level lower than 30% can cause the release of fungal spores and irritation of the eyes and upper respiratory system due to dryness [9,17]. Indoor CO_2 levels are primarily influenced by the number of occupants and ventilation rates, with sufficient ventilation being crucial for maintaining a healthy learning environment and preventing adverse health effects among children [23]. While CO_2 is not considered a pollutant indoors, it serves as an indicator of adequate ventilation levels.

The World Health Organization (WHO) sets the permissible concentration of CO_2 in enclosed spaces up to 1000 ppm. Beyond this threshold, occupants may experience headaches, nose and throat irritation, decreased concentration, and fatigue. In modern automatic ventilation systems, elevated CO_2 levels trigger ventilation processes [24–26]. According to Israeli standard 5281 for "Construction of Sustainable Buildings (Green Buildings)"—Section 5 "Health and Welfare", automatic ventilation should activate when CO_2 levels exceed 300 ppm [27]. A comparative study examining naturally ventilated classrooms in different climates (Spain and Poland) found that CO_2 concentrations exceeded recommended levels within one hour of full occupancy [25]. Another study monitoring CO_2 levels in a conference room at Bialystok University (far from industrial areas) showed a linear increase in CO_2 levels, attributed to occupants due to the supposed clean outdoor environment [28].

Indoor PM2.5 originates from various outdoor sources such as motor vehicles, power plants, dust storms, and smoking, as well as indoor sources like building occupants and materials [5]. It can travel long distances and is a leading cause of global mortality due to air pollution [6,29]. The Organization for Economic Co-operation and Development (OECD) considers PM2.5 the most abundant and critical factor affecting human health, with PM alone linked to approximately 2.1 million premature deaths annually worldwide [30–32].

Ambient PM concentration, exposure duration, individual physical characteristics (e.g., age, gender, health status), and lung development and breathing rates influence intake levels and the severity of harm [15,33]. A 2019 preschool study reported high exposure levels to PM2.5 in 69.0% of classrooms, primarily in urban areas, related to seasonal factors, occupancy, and activity patterns [23]. In efforts to protect young children during warm and cold seasons, Brazilian educational teams inadvertently sealed spaces, halting ventilation

and exposing children to harmful pollutants, particularly respiratory particles [15,34]. Epidemiological studies have linked daily PM2.5 exposure to illness and even immediate deaths, with no safe limit set by WHO [34].

In most cases, ASHRAE (the American Society of Heating, Refrigerating, and Air-Conditioning Engineers) standards are used to determine the recommended levels of each parameter screened in this study. The IMOH advises maintaining indoor thermal factors within certain ranges: a T of 20–25 °C and a RH of 30–60%. This aligns with the EPA (the U.S. Environmental Protection Agency) recommendation of keeping RH between 30% and 50% in indoor school environments, or at least below 60%, to prevent mold growth [35,36].

Our study was conducted in Eilat (Figure 1), a city uniquely positioned on the shores of the Red Sea, bordering Asia, Africa, and the Mediterranean Sea. The city's location exposes it to dust and sandstorms from Africa and aerosols from Europe, all contributing to PM levels [31,37]. Eilat's climate, characterized as hyper-arid, features warm and dry desert conditions with limited precipitation. In July–August, average daily maximum temperatures reach 39.2 °C, with relative humidity at around 16% [38–40].

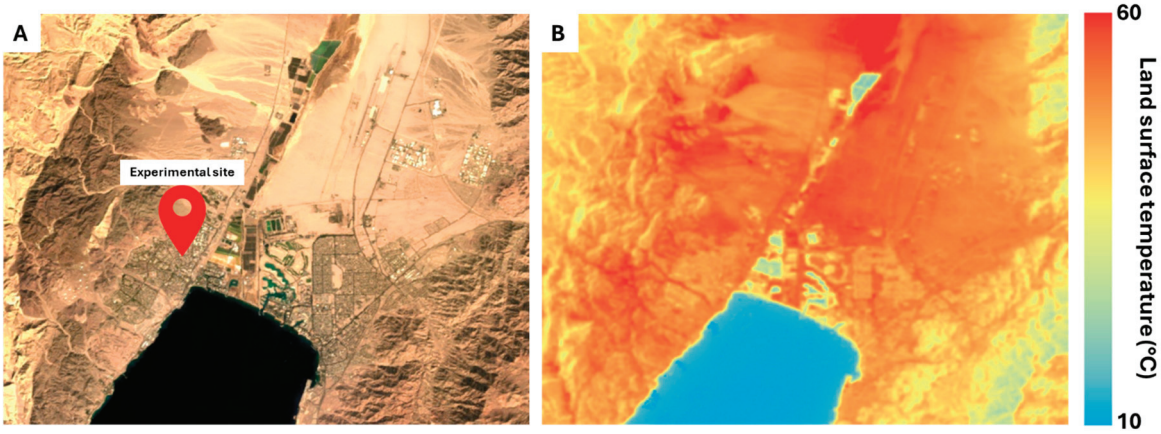


Figure 1. (**A**) A map of the general region of Eilat on the shores of the Red Sea (Landsat). (**B**) A land surface temperature map of the region of Eilat during the time of measurements (Landsat).

Desert dust storms (DDS) are natural events common in arid regions like Eilat, causing reduced visibility and haze formation due to airborne sand particles [41]. Israel, situated in the Dust Belt, experiences frequent dust storms originating from the Arabian Peninsula and North Africa, posing significant public health risks due to elevated PM levels [42]. Sandstorms raise PM levels and can cause respiratory diseases. A study examining the impact of sandstorms on air pollution and health in our region found that PM levels grew exceedingly higher than the WHO recommendation due to these events. The study found a causal relationship between an increase in PM10 levels and a rise in hospitalizations for respiratory problems [43].

Despite numerous studies highlighting the importance of indoor air quality (IAQ), real-time monitoring of IAQ factors remains underutilized. Furthermore, the significance of such real-time systems is rarely demonstrated in terms of daily variations in IAQ. These systems are particularly crucial in regions prone to extreme weather events that can significantly impact IAQ, especially in areas with limited climate control systems. In our study, we installed a sensor network in a public elementary school to monitor four key environmental factors: temperature and relative humidity (T and RH%) for comfort and carbon dioxide (CO_2) and fine particulate matter (PM2.5) for health. To highlight the critical need for such a system, this paper focuses on one month's measurements, including an extreme dust storm, and its impact on indoor environmental conditions.

2. Materials and Methods

2.1. Study Site and Data Collection Periods

Our study was conducted at Ye'elim Elementary School in Eilat. This school, constructed in 1967, is situated in a quiet urban area with low socio-economic status. The city's geographical location, bordering the Arava Valley to the south and the northern coast of the Red Sea, along with its hyper-arid climate, imparts distinctive characteristics [38] that may exacerbate indoor air quality issues at the school.

A modular monitoring system (Section 2.2) was installed at the school to enable real-time monitoring of air quality and various environmental parameters. To make this data accessible to students, a user-friendly interface was created (see Figure 2), allowing live data to be displayed for each classroom. Additionally, a color-coded display was used to help young students quickly identify which measurements were outside the recommended ranges.

Figure 2. The accessible display of monitored environmental parameters, which provides an effective way to communicate measurements to students.

Measurements of IAQ comfort parameters, temperature (T) and relative humidity (RH), and indoor environment atmospheric concentrations of CO_2 and PM2.5 were taken continuously. We chose to present measurements during the month of May 2023, which had a very big extreme weather event in the area. These assessments aimed to evaluate the levels of these parameters in unique extreme conditions characterized by an arid climate and during a sandstorm coupled with poorly ventilated spaces. We measured four classrooms (detailed in Table 1) situated on various floors of the school building, including the ground floor, second floor, and third floor adjacent to the roof. These classrooms featured varied occupancy levels, ranging from full occupancy to empty. Classrooms with distinct characteristics were treated as separate microenvironments, while others were deemed representative of the same microenvironment.

Table 1. Class characteristics: maximal number of children per class, class density, window directions, and building floor.

Class	1	2	3	4
Size (m^2)	47.5	69.0	69.0	47.5
Number of students	26	26	10	26
Students # per 100 m^2 [a]	54.7	37.7	14.5	54.7
Window direction North [b] /South [c]	South	South	North and South	South
Floor	2nd	2nd	1st (Ground)	3rd

[a] Recommended number of students per 100 m^2 (ASHRAE) is 25; [b] sun all day long; [c] shadow all day long.

The classrooms were chosen to represent various conditions and orientations. Classes 2 and 3 are situated at the corners of the buildings, while Class 4 is near the roof. Class 1 is positioned in the center of the second floor. Other factors influencing IAQ parameters analyzed in this study include the building floor on which a class is located, temperature, and humidity patterns [16,17]. In Class 3, designated as special needs classes, the student density was lower than the maximum recommended density of 25 students per 100 m^2 according to the ASHRAE standard. In contrast, in the remaining four classes, this number significantly exceeded recommendations [23].

A comparative analysis was conducted by juxtaposing the indoor data with corresponding information obtained from a monitoring station to assess the indoor–outdoor coupling (I/O ratio). This Ministry of Environmental monitoring station, measuring PM10, is situated in another district of Eilat called the "Shachmon neighborhood" (coordinates: 34°56.08′ N 29°50.32′ E) located ~3 km from the school. Data collected by this official station is accessible on their websites, enabling retrieval at any time based on the specified dates and hours.

2.2. Monitoring and Analysis Infrastructure

A modular monitoring system applied in this study was designed so that the omission or addition of components would not interfere with its overall performance. The system includes wireless control for collecting and retaining information in the cloud provided by the device manufacturers. Multifunctional instruments are used for ongoing monitoring of indoor levels of CO_2, PM2.5, and comfort parameters T and RH. The sensor's specifications are detailed in Table 2.

Table 2. Sensor specification.

Device Type (Location)	Parameter	Measurement Principles	Range
Tuya_Air (Indoor)	Temperature and Relative Humidity	solid-state sensors	Detection T: −10–50 °C
			RH: 20−85%
	CO_2	infra-red detector	CO_2: 0–5000 ppm
	PM2.5	laser scattering	NA μg/m^3

Sensors were strategically positioned in coordination with Eilat's municipality and the school's management, ensuring their approval and placement two meters above the floor. Factors such as the age of the building and the necessity to safeguard the equipment from unintentional damage were also considered during sensor placement. Measurements were conducted continuously for the entire duration of the experiment, 24 h per day. The

selected days were retrieved from this database, enabling a comprehensive analysis of indoor air quality (IAQ) parameters. Before installation on-site, the indoor sensors and communication systems underwent verification to ensure that their performance met the manufacturer's specifications, as detailed in Table 2. An interim assessment of the study sensors' measurements of T and RH at Ye'elim school was conducted against certified data to confirm their adherence to the manufacturer's specifications. Additionally, the sensors were placed together and compared to each other as well as to another device's PM2.5 measurements (Dylos DC1700, Dylos Corporation, Riverside, CA, USA). Due to minor deviations (up to 5%), the sensors were recalibrated against the Dylos device to ensure they showed similar magnitudes. The Dylos itself is not a reference device; however, the used Dylos device is constantly calibrated with PALAS Promo 2000 (Palas GmbH, Karlsruhe, Germany), a much more accurate device. The calibration was based on a linear regression between the Tuya air device (Tuya Inc., Hangzhous, China) and the Dylos device. To enhance the robustness of the measurements, the positions of each Tuya Air device were rotated and switched between classrooms.

2.3. Data Analysis

Data collected from the calibrated monitoring station (as detailed in Section 2.1) and indoor sensors were subjected to analysis. Subsequently, the average values of the environmental factors measured were compared against state-of-the-art international and national recommendations and guidelines, as delineated in Table 3.

Table 3. Recommended ranges or maximum levels of the various measured indoor parameters.

Parameter	Authority	Advised Level
T	IMOH [a]	20–25 °C
RH	IMOH [a]	30–60%
CO_2	ASHRAE [b]	1000 ppm
PM2.5	WHO [c]	5 μg/m^3 annual average 15 μg/m^3 24 h average
PM10	WHO [c]	15 μg/m^3 annual average 45 μg/m^3 24 h average

[a] Israeli Ministry of Health—https://www.health.gov.il/English/Topics/EnviroHealth/Environmental_Contaminants/Pages/intern-structural.aspx (accessed on 1 August 2024); [b] ASHRAE—American Society of Heating, Refrigerating and Air-Conditioning Engineers [44]; [c] WHO—World Health Organization [45] https://www.who.int/news-room/feature-stories/detail/what-are-the-who-air-quality-guidelines (accessed on 1 August 2024).

The study data were compared against the respective target values, which are believed to have health implications as per the Clean Air Act law. The target values for CO_2 and PM2.5 represent the maximum allowable concentration of pollutants in the air for specified durations. Anomalous results observed on a single day are considered isolated events. A deviation lasting up to four days a year is deemed within normal limits, while deviations beyond this period are considered significant and warrant attention.

3. Results

3.1. General Conditions in the School Under an Arid Climate

The values of the observations from May 2023 are presented in Figure 3. Regarding thermal comfort parameters, the measured temperatures (Figure 3A) range between 17 °C and 33 °C. During occupancy hours (8:00–14:00), temperatures are generally cooler due to the operation of local air conditioning systems. The relative humidity (Figure 3B) ranges from 20% to 60%. We observed some deviations: all the maximum temperature values and

most of the minimum values during study hours (8:00–14:00) fall outside the recommended IMOH range (Figure 3A). However, the relative humidity values are mostly within the recommended range (Figure 3B).

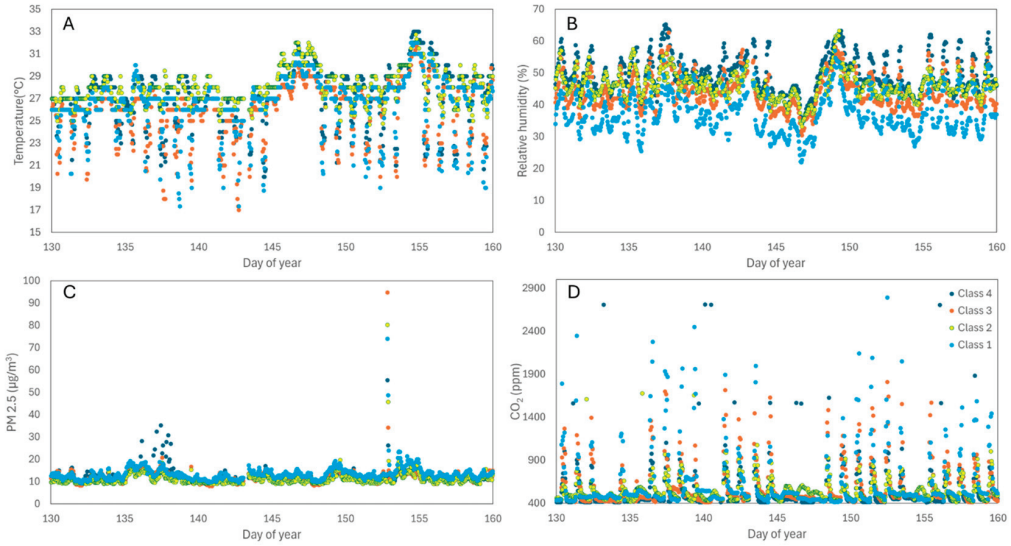


Figure 3. Observations of temperature (**A**), relative humidity (**B**), PM2.5 (**C**), and CO_2 (**D**) during 130 and 160 day of year (DOY) 2023. The observations are class-specific (four classes in total).

Interestingly, the classroom with the lowest occupancy, Class 3, had significantly higher temperatures (95% confidence level) compared to the other classrooms (Figure 4A). It should be noted that classroom differences in temperature and relative humidity can arise from various factors besides occupancy levels, such as different air conditioning systems, varying habits of opening and closing doors and windows, structural differences, and different itineraries.

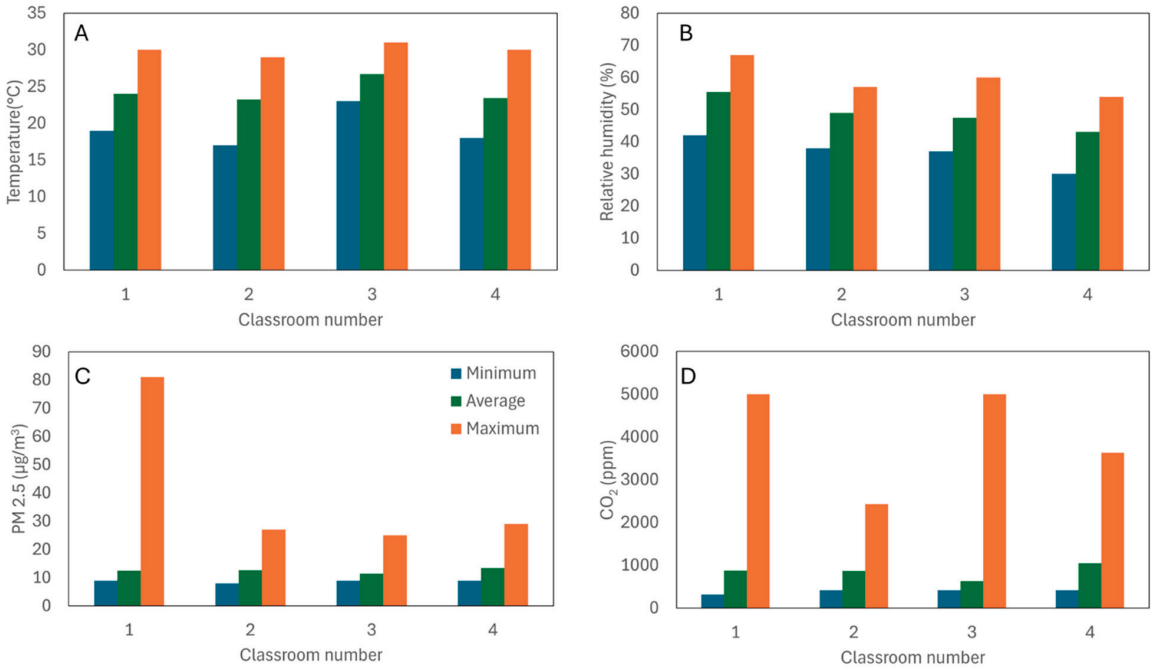


Figure 4. Daily (8:00–14:00) average, maximum and minimum values of temperature (**A**), relative humidity (**B**), PM2.5 (**C**), and CO_2 (**D**) during 130 and 160 days of year (DOY) 2023. The observations are class-specific (four classes in total). In (**D**), the maximum values of classrooms 1 and 3 reached the maximum detection level.

The PM2.5 observations revealed levels higher than the recommended WHO exposure values. This is unsurprising, as most of the world experiences PM concentrations exceeding WHO recommendations. During the 30 days of measurement, two irregular events with elevated PM concentrations were identified: days 136–139 (Event 1) and 153–155 (Event 2).

Event 1 is likely a local occurrence, as there is no evidence of it in the PM10 measurements from the Ministry of Environmental Protection monitoring station located approximately 2 km from the school (Figure 5C). In contrast, Event 2 is clearly visible in Figure 3, indicating a regional dust storm. Furthermore, the Israeli Ministry of Environmental Protection classified the days of Event 2 as extreme sandstorm days, with high temperatures, low relative humidity, and high atmospheric PM concentrations, which are deemed extremely harmful to human health.

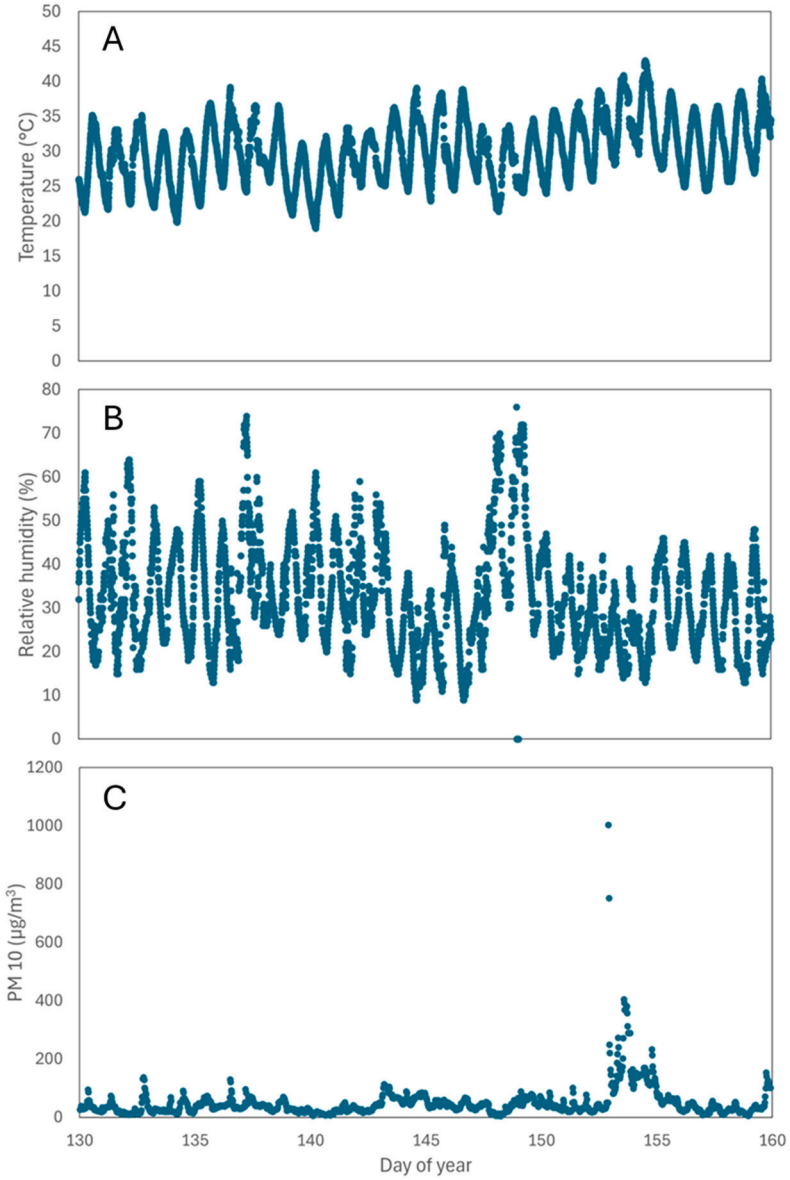


Figure 5. Corresponding temperature (**A**), relative humidity (**B**), and PM10 (**C**) measurements. Temperature and relative humidity were measured by a station of the Israeli Meteorological Service. The PM10 was measured by a monitoring station operated by the Israeli Ministry of Environmental Protection.

The PM data presented in Figure 4C shows that the minimum and average concentrations in the tracked classrooms are similar, with no significant differences found (95%

confidence level). However, the maximum values differ, with Class 1 exhibiting the highest concentrations, reaching 81 $\mu g/m^3$. Higher PM2.5 concentrations were observed when the school was vacant; however, this study concentrated on actual exposure, and thus it was not included here. There is no evident cause for the higher concentrations of PM2.5 in Class 1 compared to the other classrooms. This emphasizes the need to track air quality in real time and act to mitigate high indoor air pollution caused by outdoor events.

Unlike temperature, relative humidity, and PM2.5, which depend greatly on outdoor conditions, CO_2 concentrations mainly depend on the number of students occupying the classroom and the ventilation system. The CO_2 concentrations (Figure 3D) support this, showing higher values when the classrooms are occupied (8:00–14:00). Given that the school lacks a central ventilation system, the main reason for high CO_2 concentrations is the number of students. This is clearly shown in Figure 4D, where Class 3, which has a much lower student number per area (Table 1), produced lower CO_2 levels. All classrooms experienced higher than recommended CO_2 levels (Table 3), coinciding with the lack of a central ventilation system and the tendency to shut doors and windows due to high heat and PM concentrations.

3.2. Impact of the Sandstorm on the Indoor Environment

The impact of the sandstorm is shown in Figure 6. This event resulted in elevated PM levels and higher temperatures. First, it is evident that CO_2 concentrations, which do not substantially differ outdoors during a sandstorm, were not affected by the storm. At this time of year, temperatures are generally higher (Figure 5A), prompting students to close doors and windows regardless of the extreme event.

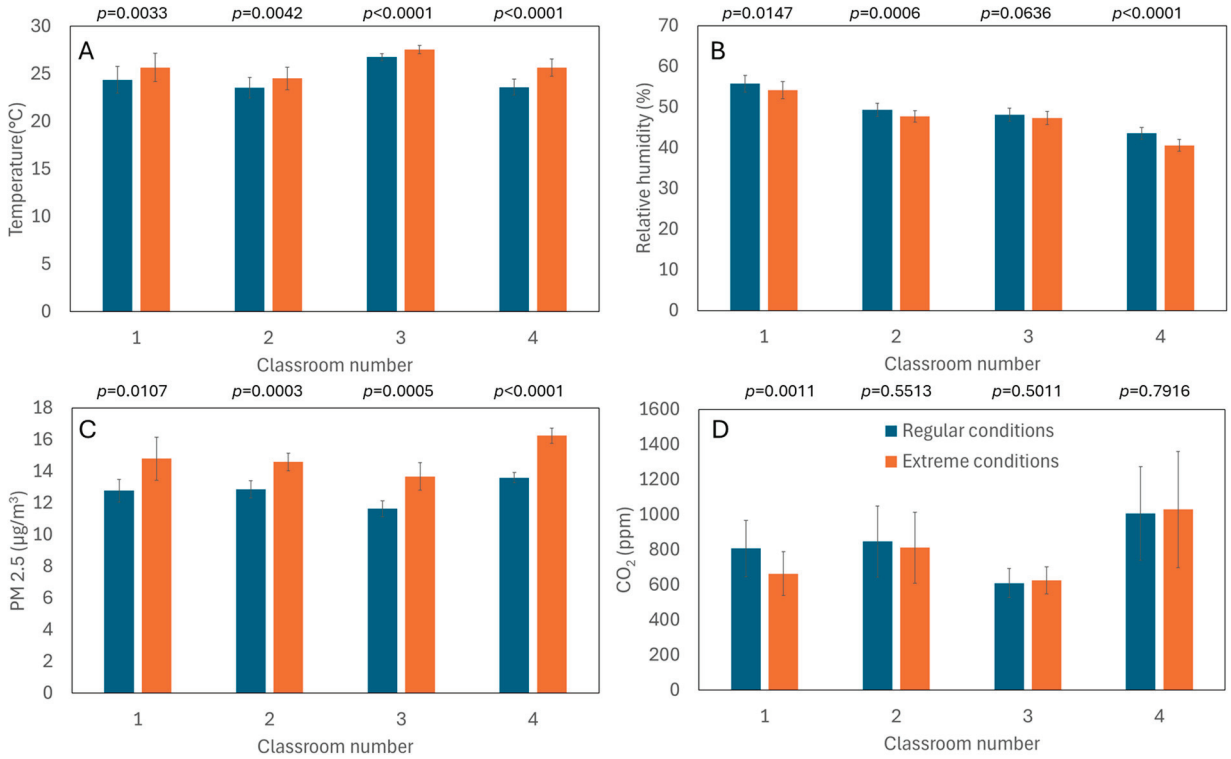


Figure 6. Daily (8:00–14:00) average values of temperature (**A**), relative humidity (**B**), PM2.5 (**C**), and CO_2 (**D**) under regular conditions and under extreme conditions. The bars represent standard deviation, and above each two bars is the *p*-value for the test of significant difference (95% confidence level).

Temperature and relative humidity were significantly different (95% confidence level) during the sandstorm compared to regular conditions. Temperatures were approximately

5% higher during the sandstorm, likely because air conditioning systems had to work against higher outdoor temperatures. This increase pushed the average classroom temperature beyond the maximum recommended level. Relative humidity values, while still within the IMOH recommended range, were slightly lower during the sandstorm, coinciding with the lower relative humidity levels measured outside (Figure 5B) and the impact of a prolonged air conditioning activity that dries the indoor air.

The most significant change caused by the sandstorm was the elevated indoor PM levels. The average daily (8:00–14:00) concentrations increased by more than 16% due to the sandstorm. These significant changes indicate a substantial outdoor air pollution event that leaked into the indoor environment.

The correlation between indoor PM2.5 and outdoor PM10 measurements is presented in Figure 7. This figure illustrates the differences in correlation between regular and extreme conditions across various classrooms. When examining these correlations, we observe that the slopes of the linear regressions are steeper under regular conditions. Although maximum outdoor PM10 concentrations during extreme conditions were approximately five times higher than those under regular conditions, indoor PM2.5 concentrations during extreme conditions were only slightly higher than those during regular conditions.

Figure 7. Indoor PM2.5 compared to outdoor PM10 at the test site. Data representing regular conditions were collected from DOY 140–150, while data representing extreme conditions was collected from DOY 153–155. The dotted line represents a linear regression based on the data.

The correlations under regular conditions are stronger compared to those under extreme conditions. However, this could be attributed to the larger dataset for regular conditions (240 data points) versus extreme conditions (60 data points). Additionally, differences between classrooms are evident in their dependence on outdoor concentrations. A higher regression slope indicates a greater dependency of indoor PM concentrations on outdoor PM levels. The fourth classroom has the steepest slope, suggesting it is more sensitive to outdoor concentrations. This classroom is the only one studied that is located on the third floor, with less interference from adjacent buildings. Conversely, the first classroom shows the lowest dependency on outdoor concentrations.

4. Discussion

4.1. General Comparison of the Measured Parameters

Eilat is situated in the heart of a desert area, regularly experiencing harsh environmental conditions, including high temperatures and low precipitation, even without the consideration of extreme weather events. These conditions force local residents to remain indoors for much of the daytime, which, while improving overall comfort, does not shield them from more severe indoor conditions compared to other regions. The average temperature and relative humidity levels, measured in Ye'elim school, are mostly within recommended ranges. However, Class 3 is an exception, with an average temperature exceeding the upper limit recommended by IMOH. Examining the differences between classrooms, we find that Class 3 has the lowest occupancy rate, which does not coincide with the higher temperatures. However, another notable feature of this classroom is its two outer walls, compared to only one outer wall in the other classrooms. This likely contributes to the higher temperatures in Class 3, as it is more exposed to heat conductance from the outdoor environment.

Maintaining thermal comfort for children during the school day can be crucial for effective learning. The average temperatures at Ye'elim School are higher than the temperatures children worldwide typically find comfortable [46]. Our data show that average hourly temperatures exceed 25 °C—considered the upper limit for thermal comfort in children—36% of the time, indicating that students at Ye'elim School are learning under suboptimal conditions.

However, several studies suggest that thermal comfort should be based on outdoor temperatures, implying that each climate has its own optimal comfort range [47–49]. Using the children-based adaptive comfort model, the calculated comfort temperature for May is approximately 26 °C [47]. In this context, classroom temperatures exceed this comfort level 30% of the time, which is still significant. It should be noted that adults' comfort temperatures are generally about 2 °C higher. Since teachers, who are responsible for operating the air conditioning, may set temperatures to their comfort level, this could further compromise the children's comfort [47].

The average CO_2 levels measured were mostly below the recommended upper threshold of 1000 ppm (ASHRAE). However, there were several instances where CO_2 concentrations spiked significantly, which should have alerted the teachers if these measures were being monitored. While CO_2 is not classified as an air pollutant, it serves as an indicator of poor ventilation [50]. A significant increase in CO_2 concentrations suggests reduced oxygen levels, which are known to diminish learning effectiveness and cause headaches [51,52]. This implies that children's performance may be impaired due to the harsh climate in Eilat, as there is no ventilation system in place, and the doors and windows are usually closed, preventing natural ventilation.

Under "regular" environmental conditions, the air quality and thermal comfort parameters at Ye'elim School partly deviate from the recommended values. While temperature, relative humidity, and CO_2 are primarily considered comfort parameters, PM is classified by the WHO as a major air pollutant. The school's PM monitoring revealed that the children's exposure to PM2.5 exceeds WHO recommendations.

In 2023, the annual average PM10 concentration, measured by the Israeli Ministry of Environmental Protection station, was 39.9 $\mu g/m^3$ (with a standard deviation of 26.8 $\mu g/m^3$), significantly higher than the WHO's recommended 15 $\mu g/m^3$. Assuming the ratio between outdoor PM10 and indoor PM2.5 remains consistent throughout the year, the corresponding annual average of indoor PM2.5 at the school would be approximately 10 $\mu g/m^3$, which is twice the WHO's recommended annual PM2.5 concentration of 5 $\mu g/m^3$. This situation necessitates that decision-makers promote the monitoring of the

air quality within schools and establish appropriate guidelines for building ventilation and filtration systems.

4.2. Indoor/Outdoor PM Correlations

The results presented in Figure 7 show a direct connection between indoor and outdoor PM concentrations. The correlations between indoor PM2.5 and outdoor PM10 were positive, with an increase in outdoor PM levels leading to a corresponding increase in indoor PM levels. Given the students' activity and the absence of significant indoor PM sources in the classrooms, the fluctuations in these correlations are most likely related to the operation of the air conditioning system and the opening or closing of windows and doors.

Figure 7 reveals two main differences: varying regression slopes between the classrooms and different slope magnitudes between regular and extreme conditions. Focusing on the regular conditions, the classroom on the highest floor (Classroom 4) exhibited the steepest regression slope (0.0795), indicating that outdoor PM had the greatest impact on indoor PM levels. Classroom 1, located on the second floor, had the lowest slope (0.0489), while another second-floor classroom (Classroom 2) showed a higher slope (0.0644).

Although both second-floor classrooms face the same direction (south), they have significantly different student population densities (54.7 vs. 37.7 students per 100 m^2). Students act as a filter for particulate matter (PM), and a higher student density theoretically leads to more efficient air filtration. Assuming a breathing rate of 0.42 m^3/h [53], the students in Classroom 1 can "filter" the entire classroom volume (118.75 m^3) in ~652 min, while in Classroom 2, with a volume of 172.5 m^3, it takes them approximately 948 min.

The classroom on the first floor exhibited a slope of 0.0634. Given the interference from adjacent buildings, a lower slope would be expected. However, this classroom has the lowest student population density (14.5 students per 100 m^2) and more windows compared to the other examined classrooms, which may account for the observed results.

The differences in regression slopes under regular and extreme conditions may indicate behavioral changes. During the apparent sandstorm, students and teachers were likely more inclined to stay indoors and close windows and doors more thoroughly. Katra and Krasnov (2020) measured PM10 levels across Beer Sheva during dust storms and found similar results, where the correlation between indoor and outdoor PM was stronger under regular conditions than under extreme conditions [54]. However, in a different measurement campaign, the same research group observed the opposite trend: a lower correlation between outdoor and indoor PM10 under regular conditions than during extreme conditions [55]. Due to the limitations of our equipment, we were unable to measure outdoor PM2.5 and, therefore, focused on comparing PM2.5 with PM10.

4.3. Sandstorm Occurrence and Their Impact on the Indoor Environment

The extreme sandstorm event was thoroughly documented across Israel, reaching peak levels of PM and temperature. The Israeli Ministry of Environmental Protection and the Ministry of Health advised the public to stay indoors to avoid exposure to harmful PM concentrations. However, its impact on the indoor environment was not investigated.

To better understand the occurrences of these windstorms, we analyzed the PM10 measurements from the Israeli Ministry of Environmental Protection monitoring station. The analysis followed the Ministry's guidelines based on a study by Ganor et al. (2009) [56]. Due to the absence of a PM2.5 measuring device at the monitoring station, our analysis was limited to PM10. Ganor et al.'s study included a comparison of PM10 and PM2.5 concentrations, adding robustness to their decision-making process, which we were unable to replicate [56].

According to the PM10 analysis, there were 16 sandstorms in 2023, with some lasting a few hours and others a few days (Figure 8). In total, these sandstorm events accounted for approximately 250 h throughout the year. These windstorms occurred year-round, except during the summer months when schools were not in session.

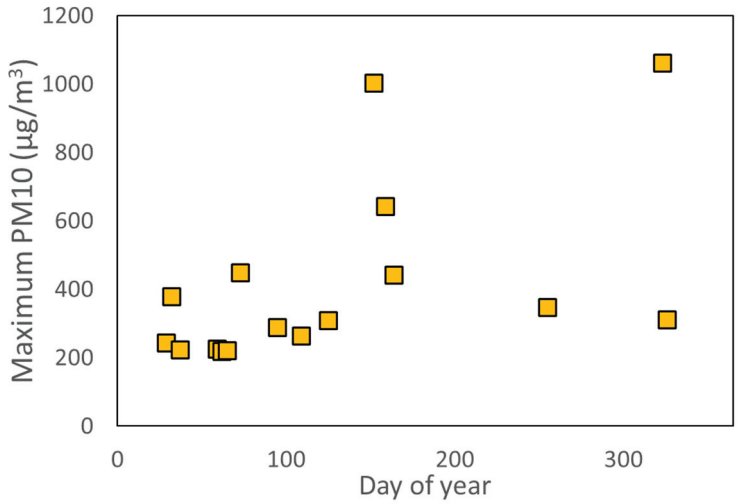


Figure 8. Maximum PM10 half-hourly concentrations during a sandstorm event. The data were taken from a monitoring station operated by the Israeli Ministry of Environmental Protection. A separate sandstorm event was defined when there was at least a 12 h difference with no sandstorms before and after the event.

This analysis (Figure 8) underscores the significance of our research, highlighting the substantial occurrence of sandstorms and their impact on the indoor environment. Our monitoring during the sandstorm event revealed notable changes, elevated temperatures (~5% on average) and increased PM concentrations (~16% on average), both of which were significantly different statistically from values under "regular" conditions. This indicates that the sandstorm was the cause of these changes. Additionally, the slight decrease in relative humidity, significantly different in all but one classroom, aligns with the drop in outdoor humidity due to the sandstorm. As expected, CO_2 concentrations remained relatively stable, as they depend primarily on the number of occupants in the classroom. The habits of opening and closing windows and doors did not change significantly during the sandstorm, likely because the high temperatures in May already encourage staying indoors with the air conditioning on.

Our study focuses on an arid region, but its conclusions are also applicable to other areas experiencing climatic changes. Many settlements in temperate regions, such as European countries, are facing rising temperatures due to climate change. While hotter periods in the past required little attention, the current temperature increases are making indoor environments, particularly schools, less thermally comfortable. These buildings often lack efficient cooling systems, similar to the situation in schools in Eilat, which negatively affects the learning experience.

Furthermore, the growing issue of desertification is expected to increase the frequency of dust storms, even in regions where they were previously uncommon. Addressing these challenges requires real-time environmental monitoring coupled with efficient indoor climate control systems to enhance climate resilience.

4.4. Research Limitations

In preparation for this research, we examined several low-cost sensors. The chosen sensor exhibited high stability, and when comparing it to reference devices, it yielded

reasonable results. However, this sensor is still a low-cost sensor, and its performance is still somewhat limited [57]. In addition, we could not place this sensor outdoors for reference due to its low durability, which left us with the PM10 measurements of the monitoring station located ~3 km from the school. The center of the research was sandstorms, a multiregional phenomenon, so the monitoring station is close enough. However, for comparison to local events, one of which apparently happened during our measurements, the 2 km distance is too far and limits our analysis in this aspect.

Another limitation is our inability to monitor the opening and closing of the doors and windows, as well as to track the occupancy levels. Filming all the classrooms was prohibited due to privacy regulations and installing open/close sensors was not applicable. Nevertheless, this information is important for the understanding of the connection between the outdoor and indoor environmental parameters.

5. Conclusions

This study investigates the importance of real-time indoor air quality (IAQ) monitoring systems, with a focus on extreme weather conditions, using a public elementary school in Eilat as a case study. A modular sensor network was deployed to measure key environmental parameters, including temperature (T), relative humidity (RH%), carbon dioxide (CO_2), and particulate matter (PM2.5). Over a one-month monitoring period, including an extreme sandstorm, data were analyzed to evaluate indoor environmental conditions and their interaction with outdoor influences.

Despite the recognized importance of IAQ, real-time monitoring remains underutilized. The sandstorm event that was measured in this research demonstrated the critical role of such systems in detecting significant IAQ deviations, including elevated PM2.5 levels and temperature spikes beyond comfort thresholds. Real-time monitoring provides actionable insights into daily IAQ variations and abnormal conditions, supporting timely interventions.

Temperature and RH analyses revealed that indoor comfort levels were often suboptimal, with temperatures exceeding recommended limits 36% of the time. Thermal comfort differences between classrooms were influenced by factors such as occupancy, building orientation, and air conditioning performance. High CO_2 concentrations during occupancy hours highlighted insufficient ventilation, emphasizing the need for improved air circulation systems.

Indoor PM2.5 levels consistently surpassed WHO guidelines, underscoring the susceptibility of indoor environments to outdoor pollution. During the sandstorm, PM2.5 concentrations increased by over 16%, reflecting substantial outdoor air infiltration. The sandstorm exacerbated indoor conditions, with heightened temperatures, reduced RH, and increased PM levels. Behavioral responses, such as closing windows and doors, partially mitigated outdoor infiltration but compounded ventilation deficiencies. These findings highlight the increased risk of poor IAQ in extreme weather, particularly in arid climates like Eilat.

The study underscores the need for targeted guidelines to improve school ventilation and air filtration, particularly in hyper-arid regions. Real-time IAQ monitoring systems can guide decision-makers in protecting vulnerable populations, such as children, from adverse health and comfort impacts. Integrating these systems with improved building designs can enhance IAQ resilience during extreme weather events, promoting sustainable indoor environments.

This should encourage schools to use monitoring sensors to track indoor environmental conditions. Proper regulations are required for the installation and maintenance of such systems to ensure a high level of accuracy. Even low-cost sensors can effectively monitor

the indoor environment if regularly maintained and calibrated. Additionally, policymakers should continue and promote studies to estimate children's exposure to harmful pollutants, establish proper regulations, and create a safer and healthier school environment.

Author Contributions: Conceptualization, E.G. and S.A.K.; methodology, E.G., S.A.K. and O.K.; validation, S.A.K. and O.K.; formal analysis, O.K.; investigation, S.A.K. and O.K.; resources, E.G.; data curation, S.A.K.; writing—original draft preparation, O.K.; writing—review and editing, O.K.; visualization, O.K.; supervision, E.G. and O.K.; funding acquisition, E.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the European Union under the ENI CBC Mediterranean Sea Basin, under the project "Berlin—Cost-effective rehabilitation of public buildings into smart and resilient nano-grids using storage" [grant ref. A_B.4.3_0034].

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors thank the Eilat Municipality and the school management and administration staff of Ye'elim School for facilitating the measurement procedures conducted within the school premises. Furthermore, appreciation is extended to government entities, namely the Israeli Ministry of Agriculture and the Israeli Ministry of Environmental Protection, for providing unrestricted access to station data, encompassing periodic datasets and associated parameters.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Choosing a path. *Nat. Clim. Change* **2023**, *13*, 873. [CrossRef]
2. Neira, M.; Erguler, K.; Ahmady-Birgani, H.; AL-Hmoud, N.D.; Fears, R.; Gogos, C.; Hobbhahn, N.; Koliou, M.; Kostrikis, L.G.; Lelieveld, J.; et al. Climate change and human health in the Eastern Mediterranean and Middle East: Literature review, research priorities and policy suggestions. *Environ. Res.* **2022**, *216*, 114537. [CrossRef] [PubMed]
3. Sun, C.; Chen, J.; Hong, S.; Zhang, Y.; Kan, H.; Zhao, Z.; Deng, F.; Zeng, X.; Sun, Y.; Qian, H.; et al. Indoor air quality and its health effects in offices and school buildings in the Yangtze River Delta. *Air Qual. Atmos. Health* **2023**, *16*, 1571–1586. [CrossRef]
4. Axelrad, R. Indoor Air Quality. *Safe Healthy Sch. Environ.* **2009**, 123–132. [CrossRef]
5. Morawska, L.; Congrong, H. Indoor Air Pollution-indoor Particles, combustion Products and Fibers. *Indoor Air Pollut.* **2018**, *73*, 37–68. [CrossRef]
6. Kurt, O.K.; Zhang, J.; Pinkerton, K.E. Pulmonary health effects of air pollution. *Curr. Opin. Pulm. Med.* **2016**, *22*, 138–143. [CrossRef]
7. CDC. *Healthy Housing Reference Manual*; US Department of Health and Human ServicesCenters for Disease Control and Prevention and U.S. Department of Housing and Urban Development: Atlanta, GA, USA, 2006; Chapter 5; p. 231.
8. Madureira, J.; Paciência, I.; Rufo, J.; Severo, M.; Ramos, E.; Barros, H.; Fernandes, E.d.O. Source apportionment of CO_2, PM_{10} and VOCs levels and health risk assessment in naturally ventilated primary schools in Porto, Portugal. *Build. Environ.* **2016**, *96*, 198–205. [CrossRef]
9. Järvi, K.; Vornanen-Winqvist, C.; Mikkola, R.; Kurnitski, J.; Salonen, H. Online Questionnaire as a Tool to Assess Symptoms and Perceived Indoor Air Quality in a School Environment. *Atmosphere* **2018**, *9*, 270. [CrossRef]
10. Argunhan, Z.; Avci, A.S. Statistical Evaluation of Indoor Air Quality Parameters in Classrooms of a University. *Adv. Meteorol.* **2018**, *2018*, 391579. [CrossRef]
11. Abhijith, K.; Kukadia, V.; Kumar, P. Investigation of air pollution mitigation measures, ventilation, and indoor air quality at three schools in London. *Atmos. Environ.* **2022**, *289*, 119303. [CrossRef]
12. Sadrizadeh, S.; Yao, R.; Yuan, F.; Awbi, H.; Bahnfleth, W.; Bi, Y.; Cao, G.; Croitoru, C.; de Dear, R.; Haghighat, F.; et al. Indoor air quality and health in schools: A critical review for developing the roadmap for the future school environment. *J. Build. Eng.* **2022**, *57*, 104908. [CrossRef]
13. Requia, W.J.; Saenger, C.C.; Cicerelli, R.E.; de Abreu, L.M.; Cruvinel, V.R. Air quality around schools and school-level academic performance in Brazil. *Atmos. Environ.* **2022**, *279*, 119125. [CrossRef]
14. Maxwell, L.E. School building condition, social climate, student attendance and academic achievement: A mediation model. *J. Environ. Psychol.* **2016**, *46*, 206–216. [CrossRef]

15. Branco, P.T.; Alvim-Ferraz, M.C.; Martins, F.G.; Ferraz, C.; Vaz, L.G.; Sousa, S.I. Impact of indoor air pollution in nursery and primary schools on childhood asthma. *Sci. Total. Environ.* **2020**, *745*, 140982. [CrossRef]
16. Kalimeri, K.K.; Saraga, D.E.; Lazaridis, V.D.; Legkas, N.A.; Missia, D.A.; Tolis, E.I.; Bartzis, J.G. Indoor air quality investigation of the school environment and estimated health risks: Two-season measurements in primary schools in Kozani, Greece. *Atmos. Pollut. Res.* **2016**, *7*, 1128–1142. [CrossRef]
17. Hall, R.; Hardin, T.; Ellis, R. *School Indoor Air Quality Best Management Practices Manual*; Washington State Department of Health: Washington, DC, USA, 2003; pp. 1–135.
18. Haverinen-Shaughnessy, U.; Shaughnessy, R.J. Effects of Classroom Ventilation Rate and Temperature on Students' Test Scores. *PLoS ONE* **2015**, *10*, e0136165. [CrossRef] [PubMed]
19. Lee, M.; Chang, S.C. Indoor and outdoor air quality investigation at schools in Hong Kong. *Chemosphere* **2000**, *41*, 109–113. [CrossRef]
20. Krawczyk, D.A.; Wadolowska, B. ScienceDirect Analysis of indoor air parameters in an Heating education building The of and Cooling Analysis indoor air parameters in an education building a using the heat demand-outdoor Assessing Dorota the feas. *Energy Procedia* **2018**, *147*, 96–103. [CrossRef]
21. Almeida, R.M.; Ramos, N.M.; de Freitas, V.P. Thermal comfort models and pupils' perception in free-running school buildings of a mild climate country. *Energy Build.* **2016**, *111*, 64–75. [CrossRef]
22. Wolkoff, P. The mystery of dry indoor air—An overview. *Environ. Int.* **2018**, *121*, 1058–1065. [CrossRef] [PubMed]
23. Branco, P.; Alvim-Ferraz, M.; Martins, F.; Sousa, S. Quantifying indoor air quality determinants in urban and rural nursery and primary schools. *Environ. Res.* **2019**, *176*, 108534. [CrossRef] [PubMed]
24. Vallecillos, L.; Borrull, A.; Marcé, R.M.; Borrull, F. Presence of emerging organic contaminants and solvents in schools using passive sampling. *Sci. Total. Environ.* **2021**, *764*, 142903. [CrossRef] [PubMed]
25. Krawczyk, D.; Rodero, A.; Gładyszewska-Fiedoruk, K.; Gajewski, A. CO_2 concentration in naturally ventilated classrooms located in different climates—Measurements and simulations. *Energy Build.* **2016**, *129*, 491–498. [CrossRef]
26. Health, W.; Regional, O. *Policies and Current Status*; WHO: Geneva, Switzerland, 2015.
27. *SI 1045*; Israeli Standard 5281 Oart 5 Construction of a Sub-Sustainable Building (Green Building). Ministry of Environmental Protection: Jerusalem, Israel, 2016.
28. Fiedoruk, T.T.K.G. The concentration of carbon dioxide in conference rooms: A simplified model and experimental verification. *Int. J. Environ. Sci. Technol.* **2019**, *16*, 8031–8040.
29. Pardo-Levin, M. *Air Pollution and Its Impact on Health*; WHO: Geneva, Switzerland, 2019; pp. 13–18.
30. OECD. *Environmental Health and Strengthening Resilience to Pandemics*; OECD: Paris, France, 2020.
31. OECD. *OECD Environmental Performance Reviews*; OECD: Jerusalem, Israel, 2023.
32. Kelly, F.J.; Fussell, J.C. Size, source and chemical composition as determinants of toxicity attributable to ambient particulate matter. *Atmos. Environ.* **2012**, *60*, 504–526. [CrossRef]
33. Gupta, T.; Agarwal, A.K.; Agarwal, R.A.; Labhsetwar, N.K. *Environmental Contaminants: Measurement, Modelling and Control*; Springer: Berlin/Heidelberg, Germany, 2014.
34. Xue, Q.; Wang, Z.; Liu, J.; Dong, J. Indoor $PM_{2.5}$ concentrations during winter in a severe cold region of China: A comparison of passive and conventional residential buildings. *Build. Environ.* **2020**, *180*, 106857. [CrossRef]
35. WHO. *DAMPNESS and MOULDS Guidelines*; WHO: Geneva, Switzerland, 2009.
36. Angelon-Gaetz, K.A.; Richardson, D.B.; Lipton, D.M.; Marshall, S.W.; Lamb, B.; LoFrese, T. The effects of building-related factors on classroom relative humidity among North Carolina schools participating in the 'Free to Breathe, Free to Teach' study. *Indoor Air* **2015**, *25*, 620–630. [CrossRef] [PubMed]
37. Zhang, S.; Broday, D.M. Predictors of the Indoor-to-Outdoor Ratio of Particle Number Concentrations in Israel. *Atmosphere* **2020**, *11*, 1074. [CrossRef]
38. Pearlmutter, D.; Erell, E.; Etzion, Y.; Yodan, R.; Meir, I.A.; Nahon, K. *Design Manual for Bio-Climatic Construction in Israel*; Ministry of Energy and Infrastructure, Research and Development Division: Jerusalem, Israel, 2010; p. 160.
39. Sofer, M.; Potchter, O. The urban heat island of a city in an arid zone: The case of Eilat, Israel. *Theor. Appl. Clim.* **2005**, *85*, 81–88. [CrossRef]
40. Saaroni, H.; Maza, E.; Ziv, B. Summer sea breeze, under suppressive synoptic forcing, in a hyper-arid city: Eilat, Israel. *Clim. Res.* **2004**, *26*, 213–220. [CrossRef]
41. Achilleos, S.; Michanikou, A.; Kouis, P.; Papatheodorou, S.I.; Panayiotou, A.G.; Kinni, P.; Mihalopoulos, N.; Kalivitis, N.; Kouvarakis, G.; Galanakis, E.; et al. Improved indoor air quality during desert dust storms: The impact of the MEDEA exposure-reduction strategies. *Sci. Total. Environ.* **2022**, *863*, 160973. [CrossRef]
42. Krasnov, H.; Katra, I.; Friger, M. Increase in dust storm related PM10 concentrations: A time series analysis of 2001–2015. *Environ. Pollut.* **2016**, *213*, 36–42. [CrossRef] [PubMed]

43. Reingewertz, Y. Article in The Israel Medical Association Journal. IMAJ 2015; 17: 628–632 [Online]. Available online: https://www.researchgate.net/publication/287260762 (accessed on 1 August 2024).
44. ASHREA. *Read-Only Versions of ASHRAE-Standards*; ASHREA: Peachtree Corners, GA, USA, 2022. Available online: https://www.ashrae.org (accessed on 1 August 2024).
45. WHO. *WHO Global Air Quality Guidelines*; WHO: Geneva, Switzerland, 2021.
46. Trebilcock, M.; Soto-Muñoz, J.; Yañez, M.; Martin, R.F.-S. The right to comfort: A field study on adaptive thermal comfort in free-running primary schools in Chile. *Build. Environ.* **2017**, *114*, 455–469. [CrossRef]
47. Teli, D.; Bourikas, L.; James, P.A.; Bahaj, A.S. Thermal Performance Evaluation of School Buildings using a Children-based Adaptive Comfort Model. *Procedia Environ. Sci.* **2017**, *38*, 844–851. [CrossRef]
48. Jindal, A. Thermal comfort study in naturally ventilated school classrooms in composite climate of India. *Build. Environ.* **2018**, *142*, 34–46. [CrossRef]
49. Heracleous, C.; Michael, A. Thermal comfort models and perception of users in free-running school buildings of East-Mediterranean region. *Energy Build.* **2020**, *215*, 109912. [CrossRef]
50. Zivelonghi, A.; Lai, M. Mitigating aerosol infection risk in school buildings: The role of natural ventilation, volume, occupancy and CO_2 monitoring. *Build. Environ.* **2021**, *204*, 108139. [CrossRef]
51. Bogdanovica, S.; Zemitis, J.; Bogdanovics, R. The Effect of CO_2 Concentration on Children's Well-Being during the Process of Learning. *Energies* **2020**, *13*, 6099. [CrossRef]
52. Wargocki, P.; Porras-Salazar, J.A.; Contreras-Espinoza, S.; Bahnfleth, W. The relationships between classroom air quality and children's performance in school. *Build. Environ.* **2020**, *173*, 106749. [CrossRef]
53. Faria, T.; Martins, V.; Correia, C.; Canha, N.; Diapouli, E.; Manousakas, M.; Eleftheriadis, K.; Almeida, S. Children's exposure and dose assessment to particulate matter in Lisbon. *Build. Environ.* **2020**, *171*, 106666. [CrossRef]
54. Katra, I.; Krasnov, H. Exposure Assessment of Indoor PM Levels During Extreme Dust Episodes. *Int. J. Environ. Res. Public Health* **2020**, *17*, 1625. [CrossRef] [PubMed]
55. Krasnov, H.; Katra, I.; Friger, M.D. Insights into Indoor/Outdoor PM Concentration Ratios due to Dust Storms in an Arid Region. *Atmosphere* **2015**, *6*, 879–890. [CrossRef]
56. Ganor, E.; Stupp, A.; Alpert, P. A method to determine the effect of mineral dust aerosols on air quality. *Atmospheric Environ.* **2009**, *43*, 5463–5468. [CrossRef]
57. Suriano, D.; Prato, M. An Investigation on the Possible Application Areas of Low-Cost PM Sensors for Air Quality Monitoring. *Sensors* **2023**, *23*, 3976. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

applied sciences

MDPI

Article

Challenges of Engineering Skillsets Essential for Driving Circularity of Smart Cities

Benjamin Neale [1], Sakdirat Kaewunruen [1,*], Dan Li [2], Ayfer Donmez Cavdar [3], Rand Askar [4], Tatjana Tambovceva [5] and Diana Bajare [6]

[1] Department of Civil Engineering, University of Birmingham, Birmingham B15 2TT, UK; bxn165@alumni.bham.ac.uk
[2] School of Urban Construction and Safety Engineering, Shanghai Institute of Technology, Shanghai 201418, China; danli1992@sit.edu.cn
[3] Department of Forest Industry Engineering, Faculty of Forestry, Karadeniz Technical University, 61080 Trabzon, Türkiye; adonmez@ktu.edu.tr
[4] Institute for Sustainability and Innovation in Structural Engineering (ISISE), Department of Civil Engineering, University of Minho, 4800-058 Guimaraes, Portugal; rand.askar@hotmail.com
[5] Governance and Security Institute, Faculty of Engineering Economics and Management, Riga Technical University, Kalnciema Street 6-506, 1048 Riga, Latvia; tatjana.tambovceva@rtu.lv
[6] Institute of Materials and Structures, Faculty of Civil Engineering, Riga Technical University, Kipsalas Street 6A, 1048 Riga, Latvia; diana.bajare@rtu.lv
* Correspondence: s.kaewunruen@bham.ac.uk; Tel.: +44-(0)-121-414-2670

Abstract: This study aims to define specific transferable engineering capabilities needed for the implementation (design and practices) of circular economy (CE) within a smart city setting. We conducted a critical literature review of over 100 studies on the core values of CE and smart cities to investigate the knowledge gap in this topic and understand what specific skillset is employed by industry experts that can be harnessed on a wider scale, which can allow for the optimization of CE. There is a lack of research on the skillsets needed to implement a circular economy in any setting, and there are very few studies on circularity practices in a smart city setting. Primary data collection allows us to bridge this knowledge gap, yielding new findings that do not already exist concerning the skillset employed by experts in the field, which can positively impact the smart city settings in which a circular economy is implemented. We conducted a qualitative analysis based on expert interviews of 21 participants who have experience in the circular economy. This information will benefit the industry by informing businesses and councils about the key skillsets and capabilities to look out for when employing people to implement any aspect of circular practices in a smart city setting, with an emphasis on enhancing efficiency, achieving deliverables, and thinking systemically to address complex challenges they may face during the implementation. We also investigated the implementation of CE in smart cities to provide a well-rounded view of the different achievements and challenges faced during the process. This mainly focuses on the work of governance in smart circular cities, a factor that has many important implications and externalities in different sectors. This study describes the methodology adopted to formulate a detailed questionnaire for expert interviews with respect to the skill gap and capabilities necessary for working in the industry, the results of which aid discussions regarding the different challenges faced in CE implementation. Our findings reveal that background knowledge in engineering and sustainability is the most 'highly critical' hard skill according to the experts, while communication and stakeholder engagement are the essential soft skills required to ensure the success of a circular economy within smart city settings.

Keywords: engineering education; technical competency; skillset; transferable capability; circular economy; smart city; built environment; expert interviews; grounded theory

1. Introduction

'Smart cities' have a famously ambiguous definition. There is no right or wrong solution, and many different researchers have attempted to craft their own definitions for smart cities based on several socio-economic and technical factors. The essence of smart cities is that they employ digital technologies such as the 'Internet of Things', which allows data to be analyzed and optimized for the system to make better informed operational decisions moving forward, with the purpose of improving several aspects of the city, which relates to the goal of being more sustainable. For example, Barcelona is a smart city; it uses information and communication technologies and governance in the economy and environment for effective urban management and is now considered a European success story of urban development [1,2]. Other cities employ such technologies and practices with the aim of replicating the successes of Barcelona, which is ranked as one of the top four best cities in Europe to start new businesses. However, employing smart technologies alone does not guarantee sustainable results. For a smart city to operate in such a way that sustainability goals are met, a circular economy (initiatives, technologies, and practices) must be introduced.

A circular economy (CE) is an economic system aimed at eliminating waste and pushing the regenerative use of resources [3,4]. The EU Circular Economy Action Plan is a manifesto promoted by the EU to prioritize circular processes and sustainable consumption to circulate resources, which reduces the demand for raw materials [5]. There are many different thought processes that support the implementation of a CE. Primarily, there is the idea that by increasing a product's lifespan, people will need to buy the item less often, or if the item is recycled, then it means that the commodities and natural resources will have to be imported and used less often. However, looking at the bigger picture, it has much more to offer, ranging from new job opportunities to the eventual reversal of biodiversity loss [6], leading to positive impacts on climate change. Extending a product's lifespan has endless benefits, such as better value for the customer (due to lower costs and because the customer will be pleased knowing that they are consciously aiming to reduce waste) as well as an unprecedented positive impact on the sustainability aims of smart cities. The issue, however, is that the current level of circularity in the EU is not sufficient to offer the profound, visible, and immediate benefits listed above, with the EU's circularity rate, i.e., the share of material resources coming from recycled waste materials, being 11.7% in 2021 [7].

Based on the critical literature review of 100+ relevant articles available on Scopus and Google Scholar databases [8], the knowledge gap analysis reveals that no articles in the literature focused on or discussed the implementation of CE within smart city settings, particularly regarding engineering skillsets and transferable capabilities or even the hard and soft skills needed to facilitate CE implementation. The keywords used for the literature review include (i) circular economy, (ii) skill, (iii) capability, (iv) smart city, and (v) Internet of Things. Boolean operators based on the words "AND" and "OR" were used in the literature search on Google Scholar (advanced search option) and Scopus. The search for the most relevant literature (overlapping with at least two keywords) resulted in over 100 articles. However, our critical review reveals the acute shortage of research on engineering skills and capability required to implement circular practices in smart cities. This has led to the following research question: what are the critical skills required to generate growth in the adoption of a circular economy for smart cities? Therefore, this study aims to investigate the fundamental skills and education needed by industry practitioners to enact circular and smart cities. Using a rigorous questionnaire that concerns the expert's hard and soft skills, language proficiency, and ability to encounter challenges and implement security and privacy measures, there is an opportunity to maximize the impact of CE on smart cities, aiming to increase circularity rates and, in turn, offer greater opportunity for innovation

and the achievement of smart city sustainability goals. Expert interviews were conducted to derive new insights. Such studies are relatively unprecedented in Europe, and the European Commission has recently funded an EU-wide action called 'CircularB' to systemically assess the challenges and barriers to circular economy implementation. By addressing the research question in this study, we can streamline our efforts to develop tools and educational programs that can overcome the challenges and barriers faced by industries. The outcome of this study will improve the adoption of CE measures in smart cities.

2. State-of-the-Art Review

Europe's CircularB Action has already identified the severe lack of circular economy implementation in built environments across Europe. In particular, existing building stocks face the most critical challenges. Smart buildings located in smart cities have been around for many decades (considered as an existing building stock), but it can be observed that many of them have not used circular economy practices to improve their sustainability. This evidence underpins our research question and supports the rationale for conducting this study. In our study, we will categorize the relevant studies into three themes: (i) smart cities; (ii) circular economy (CE); and (iii) the implementation of CE in smart cities. The literature review revealed a shortage of research and insights into CE implementation within smart cities, as detailed below.

2.1. Smart Cities

2.1.1. Internet of Things (IoT)

The Internet of Things (IoT) is a technological practice, a system that has resulted from the evolution of conventional networks and enables quick and efficient decisions. It entails a large number of sensing devices that are connected to each other [9]. These devices share and analyze data and communicate through WIFI, Bluetooth, and other means to make an informed decision; in our case, the decisions involve the operation of a smart city that is sustainable, has low emissions, and enhances the quality of services for its citizens [9]. The idea is that there is minimal human interaction when decisions are made this way, which leaves out any risk of corruption or human error, ensuring that all decisions made are backed by data to ensure the best possible outcome with respect to a smart city's sustainable development goals.

Although there are endless facts and data that are used by IoT technology in smart city decision-making, there are also drawbacks that must be considered. Ethical concerns have been raised about the General Data Protection Regulation (GDPR), security and data privacy breaches [10], and unauthorized access to connected devices; so, ethics play a big role when businesses decide to employ this technology in any capacity, with the aim of improving energy efficiency or reducing waste. Based on this, we included a key question in the questionnaire sent to industry experts: "How are privacy and security concerns due to smart city technologies in your work addressed?" Insights derived from this question will help improve our understanding of the precautions taken to mitigate and prevent these issues and will indicate other potential skills suggested by the experts, such as critical thinking, computing and software proficiency, and the indication of a growth mindset.

The Internet of Things is an 'architecture', a layered model format carefully planned by architects, with specific capabilities involving technical expertise, as shown in Figure 1. The main aim of this research is to identify the engineering capabilities and transferable skills needed to implement CE in smart cities in a way that maximizes its potential, and it has already been suggested that architects should begin to work on a new IoT framework that can optimize functionality, capabilities, and security. However, this must be performed in a way that does not cause fragmentation (disconnection of some parts of the Internet of Things from others) [11].

Studying the capabilities of industry experts may indicate the skills required to implement a revised IoT architecture that can maximize data analysis and outputs.

Figure 1. A simplified visualization of IoT architecture in smart cities.

2.1.2. Systematic Perspectives of Smart Cities

The need for smart cities has never been questioned. A substantial portion of the world's resources is utilized by cities; in fact, 75% of the total energy is consumed by cities [12]. This perpetual energy consumption generates nearly 80% of greenhouse gases, which have adverse effects on the environment [13]. This huge consumption, when managed incorrectly, can have profound adverse effects on the environment, with many expert sources predicting that as early as 2030, the damage caused by carbon dioxide emissions will be irreversible [14], which means that there will be adverse effects on the Earth's temperature in the future. Given that cities are responsible for 75% of global CO_2 emissions [15], it is imperative that CE is implemented in the correct way to combat this issue effectively.

Governments are promoting the 'smart city' solution, understanding that investments in human capital, modern ICT infrastructure, and sustainable growth are key to a functional smart city, of which all aspects can be achieved through participatory governance [16]. Section 2.3.2 provides deeper insights into governance, which is a set of policies, regulations, and institutional frameworks that guide and influence the transition to a more sustainable setup. This can include promoting social action and manifestos (such as Reduce, Reuse, and Recycle), setting quantifiable goals such as emissions caps and net zero targets for corporations, and, most importantly, adopting the 'Internet of Things' system, outlined in Section 2.1.1.

Several studies examining all socio-economic and technological aspects show that smart cities are more effective if their innovation potential and knowledge processes are at the highest standard possible [17]. The appropriate technology and infrastructure must be in place to synthesize and quantify the data for some purpose, to make the correct informed decisions based on the context, and to improve efficiency and, in turn, sustainability and quality of life [18]. To ensure that the most innovative technologies are in place and are synthesizing data, it is imperative that the people involved in its transition, design, and installation are knowledgeable about circularity practices and possess the correct engineering skillsets. This further reinforces the importance of this research on this topic.

The existing research on smart cities is insufficient as it does not focus on the real-life aspects of the implementation of smart technologies. The bulk of the research has been performed by researchers who have never implemented circularity technologies and, therefore, do not understand the different complex challenges faced in its implementation; perhaps, they also do not understand the behaviors of the public, who will be affected by the practices. The aspects of smart cities stretch far beyond smart technologies. Certain skillsets employed by experts in the field would lead to positive developments, as stated before, being transferable across all aspects of the smart city, ranging from IoT technologies to circularity schemes.

2.2. Circular Economy (CE)

2.2.1. Definition of Circular Economy

A circular economy is an economic system aimed at eliminating waste and the use of additional resources by keeping recycled materials in the economy for as long as possible, as illustrated in Figure 2. CE is 'a framework, which involves transforming how we manage resources, how we make and use products, and what we do with the materials afterwards' [6]. There are many different thought processes that support the implementation of the circular economy. There is the idea that by increasing a product's lifespan, people will need to buy the item less often, or if the item is recycled, then it means that the commodities and natural resources will have to be imported and used less often. However, in the grand scheme of things, it has much more potential, in particular, the sustainability advantages of CE being implemented correctly in a smart city setting, which is the primary focus of this study. The challenge of this study is to understand why circularity ratings are so low, especially in Europe, and to suggest the steps that can be taken to improve circularity in smart cities using industry expert questionnaires to understand the human factors that influence the efficiency of smart cities.

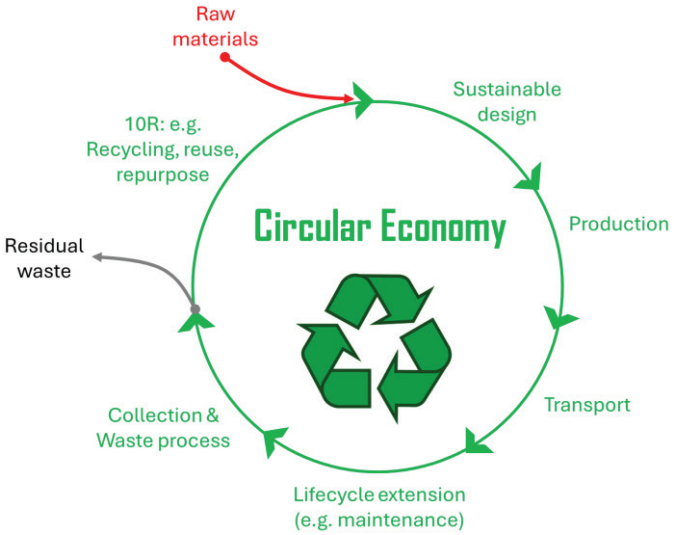


Figure 2. Infographic of a simple circular economy.

2.2.2. State of Practice of Circular Economy

In practice, there will be numerous job openings in a new sector, ranging from warehouse and floor workers to management and corporate roles. For instance, the analysis of CE-related sectors shows that Britain could have over 200,000 gross jobs created by the year 2030 [19]. CE will range from smaller day-to-day items to transport modes and potentially buildings; so, the benefits will range from more affordable and sustainable transport (as most of, if not all, methods of transport will be derived from recycled materials) and potentially cheaper construction costs as buildings will be reviewed (potentially with BIM, see Section 4 Results), aiming to reuse as much material as possible in renovations and new buildings.

CE can reduce waste and create more sustainable habits and business models [20] via the revised recycling scheme labeled as the eight Rs (refuse, reduce, reuse, refill, repair, regift, recycle, and repeat [21]) and by extending the lifecycle of products [22], which have several positive impacts on a large scale. Extending the lifespan of products results in less raw materials (and water + energy) entering the economy initially as fewer new items are being manufactured; it means that products will become cheaper as more second-hand/refurbished products will become available and that less greenhouse gas emissions are released as transportation and production are reduced.

2.2.3. Development of Circular Economy in a European Context

In 2020, the EU adopted a new plan called the Circular Economy Action Plan (CEAP), part of the European Green Deal, promoting circular processes and sustainable consumption to keep resources recirculating in the EU for as long as it is sustainable [5]. However, it is apparent that further development is needed due to deficiencies in the supply of standards, databases, or tools for the classification and monitoring of building circularity [23], which is the work undertaken by COST Actions CA21103 CircularB and partly by CA22124 ECO4All. The CircularB initiative consists of four working groups of esteemed researchers and construction industry professionals who aim to develop a common international framework for circularity rating, all harnessing their engineering skillsets in different interdisciplinary areas of research to bring together a framework based on several key performance indicators. The skillsets harnessed by these industry experts are vital as eventually, the framework in place will influence industries in smart cities to implement more circularity practices and operate in an innovative way in the future. Obtaining as many results as possible from the CircularB network would be ideal as there are few qualified practitioners to answer this study's questionnaire, and if all practitioners implementing CE in a smart city setting also harnessed these skills, the development of smart cities would surge.

2.3. Implementation of Circular Economy (CE) for Smart Cities

2.3.1. Challenges of CE Implementation

Three of the biggest struggles in implementing a CE are a lack of financial support from authorities, a lack of long-term strategic goals, and a lack of CE awareness [24]. Finances are often a concern regarding major government-backed infrastructure or schemes. Some schemes can be restrictive. For instance, historically, when governments have enforced rules that can restrict or inconvenience the public, they are often met with backlash, e.g., people protesting against the expansion of the £12.50 ULEZ charge in London [25], which means that people with cars who did not qualify would have to pay a daily charge to drive through some areas in London [26]. However, when gauging the long-term savings potential of the implementation of CE in smart cities, the immense number of positive impacts it would have in day-to-day life for the public is apparent, namely, longer product lifetimes, which in turn lower environmental impacts [27] and costs. There would be little friction from the public due to these benefits and because of people's guilty conscience, which tells them that they should live in a more sustainable and green manner; there is evidence that guilt and pride influence decisions to buy ethical and sustainable products [28].

There is a small chance of friction from individuals within the government themselves. The technological impacts have the potential to be highly beneficial, not only due to the extended life, making it more likely that people will be connected to the internet, which will vastly improve their quality of life, but also because different businesses will begin to implement more smart technology if they are guaranteed longevity (better value for money). The technology will also reduce the company's carbon footprint, which is often desired and in line with its business goals. On the other hand, it is allegedly in a politician's best interests to maximize profits for their own benefit (as well as that of their friends and 'donors' [29]) rather than save the average consumer money, which makes sense as more profits means better business performance for production companies, meaning that they receive more corporation tax; such CE will results in a significant reduction in the sales of new components and raw materials [30], which means less profits will be made, and the government will receive less tax.

Furthermore, the IoT, a context-aware network that connects a large number of sensors and devices, with the aim of sharing data and making informed decisions to positively impact the business, can be utilized. This is a huge step towards making the right business decisions

to operate in a sustainable manner, and if most businesses in a city decide to implement this technology, it would result in major decreases in energy consumption and greenhouse gases such as CO_2 (over 2.0 Gt of CO_2 emissions reduced in the energy sector by 2020 [31]) and would kickstart the process of reversing the effects of years of pollution. Although there are great benefits for businesses wanting to save money and act in a greener way, it would lead to fewer profits for the government, which prevents companies from having free rein over the implementation and could potentially result in implementation with government regulation and governance. However, if there is motivation, such as a council aiming to narrow their emission gap and contribute to sustainability, especially that of supplier companies, there would be no reason for them to oppose the implementation of circularity [32].

The skills and capabilities of those implementing CE are the main factors in approaching and solving the challenges of a lack of long-term strategic goals and a lack of CE awareness.

2.3.2. Governance of Circular Economy in Smart Cities

Governance connects the key categories of smart transition through policies, regulations, and institutional frameworks that guide and influence the transition to a more sustainable and circular way of managing resources and materials. Figure 3 displays a way of envisioning the different modes of governance and how they can ensure all aspects of a smart city work together to achieve sustainability [32]. These policies, regulations, and institutional frameworks all play a huge role in the implementation and transitions involved as they quantify the aim; quotas, goals, business models, and incentives can be set, which will encourage businesses to adopt a more circular practice, but it also means that governments can enforce these quotas and penalize businesses that choose not to change their operations (such as the emission penalties in China), with the aim of implementing circularity practices. Such policies should be considered in Europe, with the aim of increasing the circularity rate, as it was found in China that the policy had a significant greenhouse gas reduction effect, specifically on industrial SO_2 emissions [33]. These policies would have to be drafted, proposed, and implemented to be as impactful as possible; so, research and communication skillsets must be utilized between the government and all stakeholders, once more emphasizing the importance of this study, which aims to find the optimal skillset required to implement CE to the best possible standard.

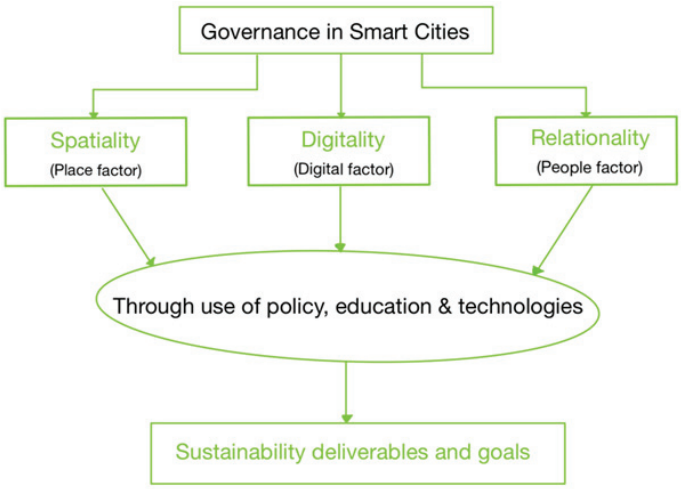


Figure 3. Visual representation of governance in smart cities.

From Figure 3, it can be observed that governance branches into three different subcategories: relationality (People), spatiality (Place), and digitality (Technology). Relationality is a mode of governance concerning relationships: relational responsibility building through transformative communication [34]. This is quite a broad mode when considering

CE; however, it mainly includes having facilities in place to educate citizens on behaviors that promote circularity practices and to reward people who make changes to act in a more sustainable manner. It also potentially makes it easier/cheaper for people to start up a small business in the CE field to accelerate the growth of the market. This also relates to engineering capabilities and transferable skills. One of the biggest challenges of CE implementation is a lack of social inclusion and participation [35]. Utilizing relationality, the people in charge of the educational facilities will pass on behaviors and advice to the public. This means that if the people who are teaching and informing others have certain skills and behaviors, we will eventually see these behaviors being displayed, and therefore, it is important that people with the correct skillsets oversee the implementation of CE.

Spatially, it is a government's ability to promote sustainable infrastructure. "Spatiality as an essential aspect of 'the urban' has a significant role in anchoring circularity through designated areas or enclaves that can be used in scaling-up CE solutions." [32]. By optimizing a smart city for circularity practice, CE activities will become more efficient; for example, having sufficient recycling facilities in place will streamline the process, and using land sustainably and having sustainable transport options (including the promotion of cycling lanes) will make it easier for people to make life choices with CE in mind.

Digitality is the mode of governance that relates to the optimization of technology, with the aim of making CE more accessible and integrated with day-to-day life, pertaining to the question of how the city government utilizes new technologies to promote CE [32]. For example, the government may decide to promote the 'Internet of Things' technologies in schemes or incentives, making it easier and cheaper for businesses to set up the technology, or maybe the government could subsidize certain software that could be used to make more streamlined and energy-efficient business decisions. It also means that the government could introduce more strict regulations on new technologies implemented, like a cap on energy usage and carbon emissions and, potentially, new guidelines on the disposal of products that have reached the end of their lives. All these steps are controlled by the government (usually in association with innovative and skilled engineers) in order to promote CE in smart cities.

Governance is a key factor in the inclusion of engineering skillsets in CE implementation in smart cities; such practices, including schemes and education, must be collaboratively researched, drawn up, and delivered by engineers. Digitality requires a particular software and computing-based engineering skillset to make the technology operational and to continually maintain and update the systems to create a smart system that is as energy-efficient and decarbonized as possible.

2.3.3. Impacts: Jobs, Environment, and Ethics

Circular economy, the practice of extending a product's regenerative life cycle via means such as the eight or ten Rs and other government schemes, reducing the need to import raw materials, means that smart city goals such as a prospering economy and improved environmental aspects will be achieved more easily by reducing the use of natural resources and enhancing knowledge about its implementation [22]. The positive impacts on the environment are endless, ranging from lower emissions to better land use, 'deconstruction', and the reuse of building materials for a less destructive impact on the urban environment [36].

Advantages can come from extending a product's life for a much longer period, such as better value for the customer (due to lower costs and because the customer will be pleased knowing that they are consciously prioritizing waste reduction and management), and it will have an unprecedented positive impact on the aims and tasks of smart city authorities [37], i.e., reducing manufacturing emissions and other deliverables. For instance,

if most manufactured items that are used daily in cities, ranging from cars and bikes to coffee cups and water bottles, had their lifecycles extended, it could result in great externalities, such as improved well-being (human health and state of mind [38]) since more people would have access to different transport modes and due to cheaper day-to-day costs, coupled with the peace of mind obtained from the fact that they are choosing to make changes to live more sustainably. It would also have the potential to create many jobs in this sector, ranging from project and process management to marketing and warehouse workers, who will have jobs such as recycling processors or machine operators.

2.4. Knowledge Gap Analysis

To address our research question (which engineering and technical skillsets are needed by industries to implement circular economy practices in a smart city setting?), we delve into the literature to define the limitations, scopes, research methods, assumptions, system boundaries, and stakeholders. This was conducted to serve our research goal and help us to design the questions for industry experts.

Evidently, no research exists on the technical challenge related to engineering skillsets required to implement CE and practices in a smart city setting, with the aim of making the system as efficient and innovative as possible; so, there was a knowledge gap in the research. Ranging from soft skills concerning communication, management, and professionalism to hard skills such as computer program proficiency and proficiency in several languages, skillsets cover a broad spectrum in engineering, which is why they are often unheeded when considering the implementation of new technologies and systems and when taking approaches to address the main challenges of CE implementation.

To identify the engineering skillsets that are needed to not only implement CE but also to maximize the efficiency of the system, with the aim of achieving a decarbonized system, the focus of this study was targeted at industry experts, i.e., individuals who have worked in sustainability and on circularity developments in a smart city setting. This was performed using an ethically reviewed questionnaire to outline the skillsets needed to work in the field, along with finding out additional details about the expert's job roles and responsibilities. The details of the questionnaire are outlined in Section 3.1.2. The implications of this oversight will often result in impacted project successes as it has been found that a worker's personal competencies (such as skills, training, knowledge, etc.) have a direct influence on the success of the project [39] and its deliverables, leading to a missed opportunity to implement CE and its technologies at the highest standard possible. Our review also found that qualitative research using the semi-structured interview method is widely used to address the topic of our research question.

The importance of this study is vital. The built environment has never been smarter, with more smart technologies harnessing the IoT than ever before to make informed decisions to help minimize energy consumption [40]; so, there is great potential for implementing CE practices to a high standard.

3. Materials and Methods

Our study was predominantly inspired by CircularB's expert group meetings. The initial insight was derived from expert meetings. Then, our research question was formulated with the knowledge gap analysis. In this study, a qualitative research method (using snowballing expert interviews) was adopted. We first identified the data and insights related to the research question. The insights helped us formulate the questionnaires or open-ended questions used to interview industry experts and stakeholders. Qualitative analysis was used to develop aggregate understandings about the trends, priorities, and significance

of engineering skills and education underpinning the factors critical to overcoming any challenge and barrier to circular economy implementation in smart cities.

3.1. Data

The method used to obtain primary data for this research is rigorous. The participant sample is an array of CircularB working group participants, professors, workers, industry professionals, and researchers with knowledge of CE and circularity implementation. This means that the results were not skewed because all participants were knowledgeable about the subject at hand and that all engineering skills and capabilities proposed will be confirmative and relevant to circularity implementation. The expert interviews were conducted in compliance with GDPR regulations and with signed consent forms. The questionnaire was subjected to a rigorous ethical review process as part of the project proposal stage to ensure that all data obtained would be used in an ethical manner and to outline how to eliminate any ethical concerns.

3.1.1. Research Data Collection

There is no previous research on the engineering skillset needed to work in the field of circularity implementation in a smart city setting. Therefore, primary data must be obtained using interviews. We formulated a questionnaire that was subjected to a full ethical review to gain background information on the industry experts, such as their qualifications and experience, the biggest challenges they have faced, how privacy and security concerns are addressed, and the specific hard and soft skills they need to perform their work.

One participant interviewed was the Head of Innovation and Partnerships in the UoB estates office, a participant heavily involved in the university's project with SIEMENS [41] in a strategic partnership to create the 'smartest' campus in the world. This scheme is an implementation of circular technologies across the campus buildings, capturing data from the estates' infrastructure and energy plants and using it for innovation, R&D activities, and teaching [41]. Because the participant is so heavily involved in the implementation, they were an ideal subject for the research. The bulk of the participants who answered the questionnaire are industry experts from the 'CircularB' and 'ECO4All' networks. These participants are professors, researchers, and esteemed personnel in the fields of sustainability and circularity practices and are thus very well-equipped to answer the questions. The questions are rigorous, open-ended, and indicative of the skillsets employed by the experts. There is a range of experts from CircularB across Europe who completed the questionnaire, offering a great insight into their transferable engineering capabilities, such as language proficiency and stakeholder collaboration, which vary depending on the country of work. An appropriate sample size was determined through a sample size derivation justified upon the review of many qualitative studies (see Section 3.1.2).

3.1.2. Questionnaire Formulation and Analysis

The purpose of this research is to bridge the knowledge gap identified after reviewing 100+ articles on this topic [8]: What skillsets are needed for the implementation of circular economy in a smart city setting, and what skillsets will produce the best results in practice? On this basis, the semi-structured interviews were preceded by a critical literature review and a desktop study. The following questions aim to attain some background information on the experts, such as their roles and responsibilities, to explore the implementation of CE in smart cities, and to determine the key factors in addressing the knowledge gap, i.e., the engineering skillsets required to perform their jobs. Stakeholder collaboration is a necessity in the development of smart cities; so, information about the involvement of stakeholders will offer insights into the interests of different stakeholders and their input on the project, with the chance of exploring challenges and disputes between engineers and stakeholders

and the skills needed to navigate these relationships. The questions concern the areas of skills and experience (E), circular economy (CE), smart cities (SC), mindset indicators (M), and engineering knowledge (K). The questionnaire questions and areas that they target are tabulated in Table 1.

Table 1. Examples of interview questions designed to gather necessary information and insights from the expert interviews.

Question	Area *
What is your role relating to a smart city or circular project/practice?	E, CE, SC, and K
What are the specific responsibilities of your role?	E
Have there been any challenges which you have faced which have hindered your ability to meet certain goals during the implementation of smart technologies?	E and SC
Have new technologies had an impact on your day-to-day work? If so, what technologies and what impacts?	E and SC
Do you believe that your role could ever fully be rendered obsolete due to the 'Internet of Things (IoTs)' or other smart technologies? (Yes/No)	SC
Is there any restriction or enforcement by governmental policy to implement circular practices or economy within the smart city framework?	E, CE, and SC
How do you involve various stakeholders in your smart city project(s)?	E, SC, and M
What do you believe the next steps will be for your business based on emerging circular technologies and practices?	CE and M
Would you like to share any other perspectives on this topic? If deemed relevant to the research these may be used as expert opinions in the dissertation and you will be duly credited.	E
Have you implemented circular economy/practices in your project(s)?	E and CE
In a nutshell, what is the main barrier to the implementation of CE that you have seen in your project(s)?	E, CE, and M
How do you stay updated on the latest developments or updates in the smart technology field relevant to your job?	SC and M
How are privacy and security concerns due to smart city technologies in your work addressed?	SC, M, and K
What role do data analytics and insights from smart technologies play in decision-making processes within your team or organization?	E, SC, M, and K
What educational background is needed for your role?	E and K
What hard skills are needed?	E and K
What soft skills are needed?	E and K
Are there opportunities for professional development?	M
Are there any employee testimonials I can have access to?	E
What skillset is needed to perform your work? Knowledge of any specific disciplines or programs?	E and K
Has the skillset required to perform your job changed since you were first employed?	E and K
Is proficiency in another language preferable for your work?	E and M
How do you apply the notions of life cycle assessments to the decision-making process in a smart city framework?	E, CE, SC, and K

* skills and experience (E), circular economy (CE), smart cities (SC), mindset indicators (M), and engineering knowledge (K).

A sample size determination was completed before the questionnaires were conducted to ensure that the results would have great statistical power, meaning that the results

would indicate that an effect is present (the effect being common transferable skills and capabilities among the industry experts working in the field). This ensures that the results obtained are as repeatable and reliable as possible whilst keeping this study efficient as a sample size that is too large would be inefficient and time-consuming. There are no strict guidelines in research for sample size determination concerning qualitative data. Rather the goal is to achieve a point of data saturation, where the data are repeated to the point where any conclusions are proven without reasonable doubt and further data collection is deemed unnecessary [42]. The most appropriate procedure was to review existing research articles and books concerning the collection and analysis of qualitative data to deduce an appropriate sample size. This is technically a grounded theory study, one that does not have a hypothesis before data are collected; thus, new insights can be derived from data [43,44]. Several research articles online state that grounded theory studies have an average sample size of 20 participants [45], a number that will result in accurate conclusions. One study states that a sufficient sample size for data saturation is 16–24 interviews [46], and two others state that a sample size of 20–30 interviews is recommended [47,48]. A qualitative study [49] on social desirability bias observed a 'Cronbach's alpha' (an indicator of consistency in results) of 0.7 when the number exceeded 12 interviews, which indicated good consistency. A larger number of questions will offer a diverse range of challenges that experts face when implementing CE and will likely indicate a clear effect of the skillsets needed to overcome said challenges.

3.2. Methods

The qualitative data underwent a thematic analysis, which is the process of identifying and analyzing a pattern or effect in qualitative data [50]. Our critical literature review also reveals that the semi-structured interview is the most suitable for our research goal. Finding underlying themes or patterns in the data is crucial to identifying the knowledge gap: what skillset is needed for the implementation of CE in a smart city setting, and what skillset will produce the best results in practice? These can be deduced by the other questions in the questionnaire. Two of the biggest barriers to CE include difficulty setting up an effective circular supply chain and challenges with customer behavior and product redesign [51]. Three questions that are particularly important in resolving these challenges are 'How are privacy and security concerns due to smart city technologies in your work addressed?', 'Is proficiency in another language preferable for your work?', and 'How do you involve various stakeholders in your smart city project(s)?'. These three questions are indicators of skillsets that are often overlooked: first, the privacy concerns would indicate hard engineering skills such as network system knowledge, online security, and continuous learning. These skills are vital in the design of a circular economy and its technologies, meaning challenges with redesign will involve problem solving and critical thinking; therefore, understanding what steps industry experts take to combat security concerns will indicate the steps needed to tackle such a big challenge in the CE discipline. Language proficiency is a hard skill that is often overlooked because it is a common language skill in global science [52]; so, proficiency in other languages will likely indicate better soft skills, such as group communication and scientific discussion. Routine transparency and communication with stakeholders (involving customers) will also help to combat one of the biggest issues faced, namely, the challenges related to customer behavior. A company's proactive efforts to communicate with customers lead to a positive relationship between them [44]. Observing the communication skills displayed by industry experts with their project stakeholders will improve relationships when implementing CE in smart city projects, meaning fewer conflicts and issues will arise. It will also help to inform the stakeholders of the intricacies of the projects being implemented, which will help to

maximize the project's efficiency given that a lack of knowledge is one of the main barriers to CE implementation [40].

4. Results

To identify critical capabilities to enrich the circular economy towards smart cities, we conducted 21 expert interviews with experienced stakeholders across the value chain. Based on the expert interviews, new insights were obtained based on the contour levels (as shown in Table 2) to demonstrate the criticality of engineering skills required to adopt circularity practices within smart cities.

Table 2. Contour levels of expert opinions with respect to the significance of engineering skills for implementing circular economy in smart cities. [Note: Green implies low risk and low impact; Yellow implies medium risk and medium impact; Purple implies high risk and medium impact; and Red implies high risk and high impact].

Listed by % of Participants	Criticality
0–25%	Low
25–50%	Medium
50–75%	High
75–100%	Very High

Expert stakeholders were asked to provide their opinions related to the essential hard engineering skills required for implementing CE in smart cities. Figure 4 portrays the expert opinions about the criticality and influence of relevant hard skills. The majority of stakeholders stress the importance of engineering and sustainability knowledge (71% of stakeholders) but do not understand the importance of research or building information modeling (BIM).

Figure 4. Data concerning the most common hard skills employed by the experts.

All experts were consulted about the importance and influence of soft skills. However, only 20 experts responded to discussions with respect to stakeholder collaboration. The results of expert interviews related to soft skills for implementing CE in smart cities are portrayed in Figure 5. It is apparent that the expert opinions are consistent and convergent. It is important to note that the majority of experts stress the importance of communication skills.

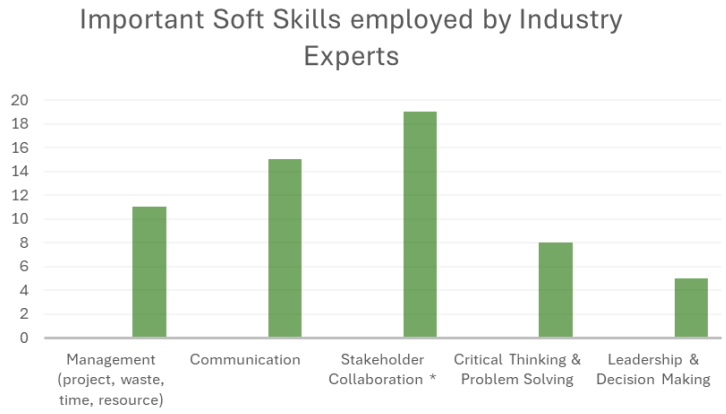


Figure 5. Data concerning the most common soft skills employed by the experts. * Only 20 participants answered the questions related to stakeholder collaboration.

Based on the expert interviews, the criticality and influence of multi-language proficiency, privacy, and security concerns are illustrated in Figures 6 and 7, respectively. The expert opinions converge on the preference for multilingual proficiency and the considerations of privacy and security concerns within smart city environments.

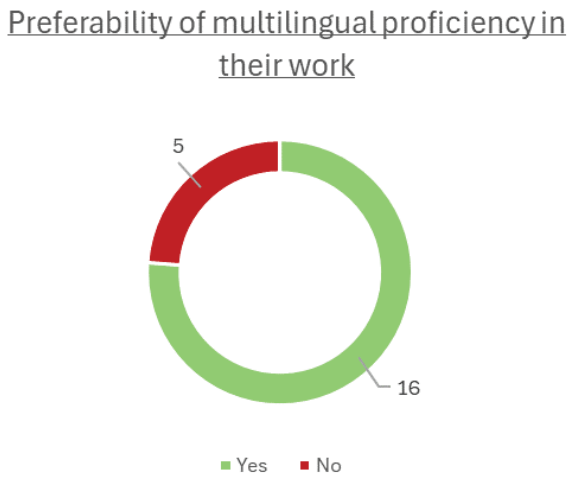


Figure 6. Expert opinions on the preferability of being multilingual in the expert's industry.

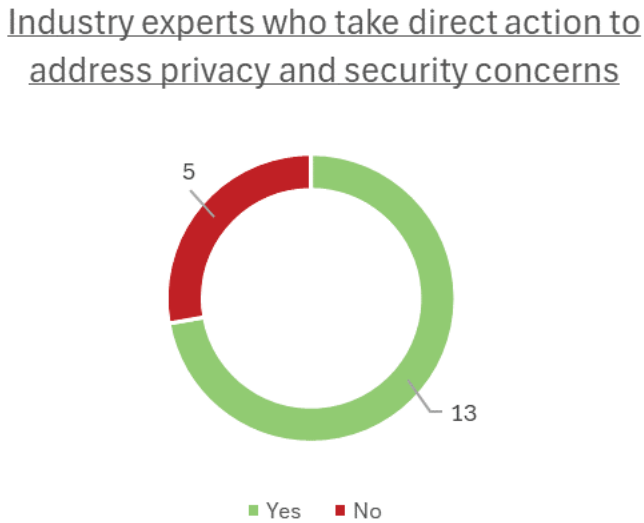


Figure 7. Breakdown of experts who take direct action to address privacy and security concerns based on the responses of 18 participants.

5. Discussion

5.1. Most Common Hard Skills

A background in engineering and sustainability is the most common hard skill utilized by industry experts, with 71.43% of them outlining the importance of the knowledge concerning their work (see Figure 4), ranking them as highly critical. These data correlate with the research of Sharif University of Technology and the University of Calgary, whose work states that engineering technical knowledge is one of the top factors that positively influence client satisfaction in the delivery of projects [53]. Students with engineering degrees bring a broad range of diverse and highly desirable skills to the market [54], and all these technical and human skills are significantly related to project success [55]. Table 3 presents the most popular results of the participants who answered the question "What educational background is needed for your role?".

Table 3. Breakdown of interviewees' expertise.

Discipline	Number of Participant Experts
General engineering	6
A degree (unspecified)	6
Civil engineering	4
Computer sciences and machine learning	4
Environmental science and engineering	4
Construction	2
Architecture	2

Around 20 participants answered this question; 11 are senior researchers and 3 are professors in universities across Europe. This means that at least 14 participants have completed a master's degree in a related discipline and are working towards or have obtained a doctorate in a related field. This also means that they will possess the transferable capabilities needed to maximize the success of the projects they deliver. The most popular solution for the most appropriate educational background outlined by the experts was that an unspecified degree was needed, tied with general engineering education, as stated by six experts. These were closely followed by environmental engineering, civil engineering, and comp science and machine learning, each one outlined by four experts. This further reiterates the importance of the skillsets employed by engineers working in the circular economy; with these capabilities and skills having a direct influence on project success, the people involved in the implementation of the circular projects must display such traits to maximize their output in smart city settings where such knowledge skills and talents are sought after [56]. Other disciplines outlined include construction knowledge, architecture, business administration, and various other outlying areas of engineering.

The next most common hard skills employed by the experts were software, computing, and coding, which were rated as highly critical, at 57.14%, followed closely by data analysis and data modeling, also rated as highly critical, at 52.38%. As stated in Section 2.3.2 (Governance of Circular Economy in Smart Cities), digitality is a key aspect of how smart cities operate, which relates to the optimization of technology in a smart city setting. Implementing circular technologies such as the IoT requires architecture, systems, and frameworks implemented and optimized by architects to support data, processing, and endpoint functions [57]. If these skills, which are desired by the experts, were more widely adopted, the current IoT architecture would be redesigned, with the aim of improving and

innovating systems, meaning the implementation of CE in smart city settings would have a greater effect.

Security and privacy concerns are another key area of digitality that concerns software, computing, and coding proficiency. When asked "How are Privacy and Security concerns due to smart city technologies addressed in your work?", 72.22% of the participants said that they take direct action when addressing privacy and security concerns due to smart technologies, of which some actions can be seen in Figure 8. This is indicative of hard skills, such as data analysis and IT, and soft skills, such as risk assessment, communication, and project management [58]. Harnessing these skills will allow CE to be implemented in smart cities in the best way possible as robust security measures will mitigate risks of serious data privacy breaches and disruption of operations, which could occur if the devices harnessing the IoT are compromised, possibly leading to financial losses, user impersonation, and server impersonation [59]. All 27.78% of participants who did not show these concerns currently work in academia. This suggests that these experts do not experience the implementation of circularity practices firsthand but rather have an academic role in the research and development of particular areas of CE.

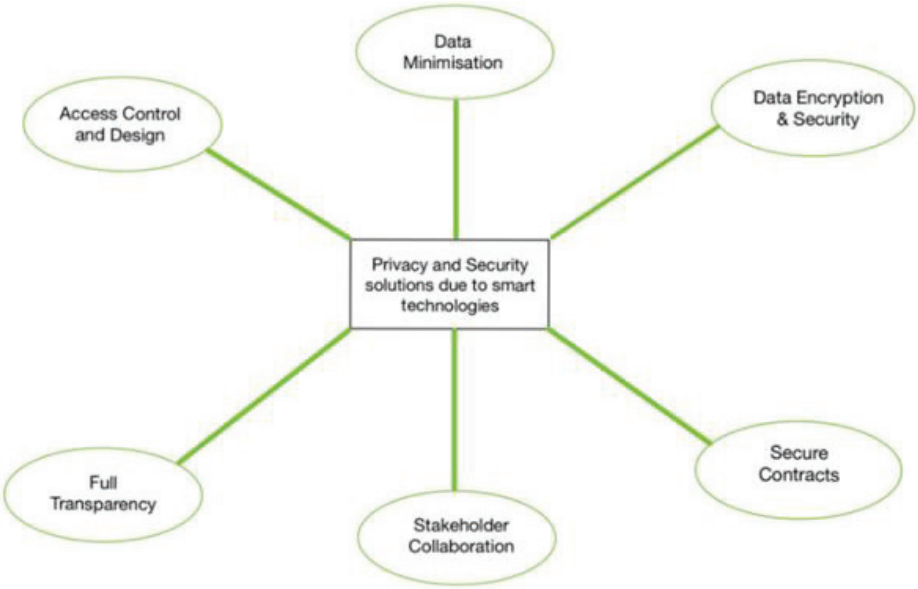


Figure 8. Approaches taken by the experts to tackle privacy and security concerns.

A total of 23.81% of the industry experts who participated in the questionnaire said that knowledge of BIM is a valuable hard skill for working in the industry. BIM is a modeling software used to demonstrate key characteristics of buildings and can simulate stages of construction over a certain time scale while providing information regarding materials, bills of quantities, energy, emissions, and other LCA information [60]. If these insights were readily available to someone with proficiency in BIM in a smart city setting, the CE could be improved as they offer an opportunity for various sustainability analyses [61], and at the end of the building's life, BIM can be used to obtain information about the materials and foundations, which can be used to reuse as many materials as possible or to observe if the renovation is possible. Tied with BIM, research skills were also outlined by 23.81% of industry experts. Research skills are required for technical development and gathering data for analysis by computers [62], as well as for understanding and quantifying the data and taking actions to improve the circularity aspects based on the data. This ties in nicely with BIM as workers proficient in both research and BIM can oversee entire circular projects: understanding sustainability data from BIM, utilizing these data, and then making informed decisions on what can be improved using research skills.

5.2. Most Common Soft Skills

Stakeholder collaboration is a key aspect of engineering, which includes liaising with investors, councils, or the community (all of whom may have conflicting interests) to determine a solution that can simultaneously satisfy multiple objectives [63]. Collaborating with stakeholders must be navigated in a way that facilitates trust. As shown in Figure 9, this is one of the key aspects of delivering a project as it has been found that greater trust between the project manager and stakeholders achieves better results [64]. A total of 95% of the industry experts who answered the question "How do you involve various stakeholders in your smart city project(s)?" said that they liaise directly with stakeholders, rating the skill as very highly critical, with five experts conducting workshops for better collaboration, five ensuring full transparency and discussions, and others employing methods such as centralized information and involving them from the design phase to delivery. The positive link between the project manager and stakeholder's trust and the result of the project emphasizes the importance of the right capabilities of the people implementing circular projects. This is indicative that a lack of collaboration and trust between managers and stakeholders is holding back the progress and effectiveness of smart cities in Europe, and if the skills utilized by the industry experts (95% stakeholder collaboration) were more widespread, there may likely be improvements in the delivery and operations of circular projects in smart city settings.

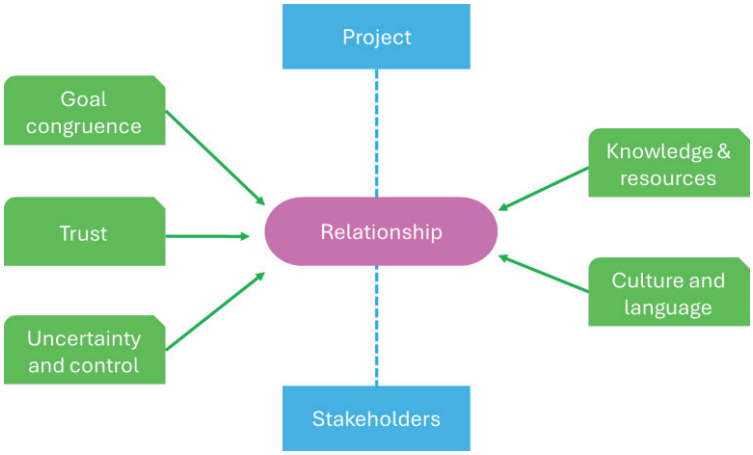


Figure 9. Trust is one of the key aspects of the relationship between project managers (PMs) and stakeholders. (adapted from [64]).

Communication is highly critical, ranked as the second most popular soft skill among the questionnaire participants, with 71.43% of the experts outlining its importance. There are several aspects of communication that influence the implementation of CE in smart cities. In response to the question "Is proficiency in another language preferable for your work?", 76.19% of participants answered that it is preferable. English is accepted as the common language for global science [52]; so, scientific discussion is conducted in English. Scientific thinking and discussion is classed as a soft skill, meaning the experts possess social and collaborative processes central to these ideas [65]. The ability to learn a second language is often coupled with higher learning capacity [66], which is a very attractive trait. It indicates that the individual will be better able to solve complex problems that will often be encountered in the digitality aspect of the CE, such as the optimization and installation of the IoT. It is also indicative of continual and multi-disciplinary learning, where the individual has the capacity to acquire knowledge in many disciplines, not just in engineering but perhaps also in business, management, and sustainability subjects, which would help them to take a more well-rounded approach to their work. As well as scientific discussion, communication links

back to stakeholder collaboration, as previously stated. Effective relationships built through effective communication are vital for a project's success [67] and, therefore, the successes of circular projects are easily influenced by communication, ranging widely from project managers and firms working on large-scale frameworks and projects with many stakeholders to workers promoting schemes and educating the public on the circular practices since it has been proven that great communication skills in education have the potential to influence people's interest and attitude, promoting a better learning atmosphere and other successes [68]. This approach can be used to combat one of the biggest issues in implementing CE, which is a lack of CE awareness (see Section 2.3.1), as having the people with the most appropriate skillset for passing on knowledge will have a profound impact on the public's awareness of the subject.

Management is the third most important soft skill, according to the industry experts' answers, also ranked as highly critical, with 52.38% of experts listing it as a skill required to conduct their work. Four areas of management were outlined, namely, time management, waste management, resource management, and project management, which are all related to each other. Project management requires skills such as leadership and interpersonal skills, technical skills, and administrative skills [69]. The technical skills required in project management cover areas of waste and resource management, closely related to the most critical hard skill emphasized by the experts, which is a background in engineering and sustainability. As stated before, technical skills are significantly related to project success, meaning these skills are highly important. Time management is perhaps the most important administrative skill [70]. It is a factor that impacts the quality and success of the project, and therefore, workers in circularity practices working in a timely manner will deliver projects of high quality, which will maximize the potential of sustainability. The third area of project management is leadership and interpersonal skills. Interpersonal skills are desirable in a project manager as they can advocate for a desired state of affairs, which promotes a positive workplace culture and trust between them and clients whilst fully respecting their rights. This will help to solve one of the main issues hindering CE development, which is a lack of long-term strategic goals (see Section 2.3.1), with positive impacts in several areas, such as stakeholder collaboration and communication disciplines, where the worker's interpersonal skills will help them work towards their goals in circular smart city projects.

Critical thinking and decision-making were highlighted by 38.1% of industry experts as important soft skills required in their roles, ranked as medium on the criticality scale. The circular economy's performance stems from the quality of the decision-making process [71]. Decision-making is a very important skill to have in conjunction with other soft skills, such as scientific discussion and stakeholder collaboration, as many different stakeholders are involved in circular projects, meaning many decisions are made as a group regarding the development of the projects. In a European context, the different working groups in the CircularB and ECO4All networks work together to create the circularity grading framework, meaning cooperation and group decision-making are necessary to foster innovation.

When asked, "Are there opportunities for personal development in your role?", 100% of the participants answered yes. Such opportunities are indicative of positive traits and capabilities. Personal development is a major indicator of a growth mindset, a trait of a worker who believes that their knowledge and abilities can flourish and develop, as opposed to a fixed mindset, where a person believes their knowledge and abilities are fixed and cannot improve. People with a growth mindset harness the skill of adaptability, being able to attack new and different challenges amidst the ever-changing climate of CE, a characteristic that is highly sought after in sustainability work as it means that candidates will have greater career interest and better persistence when challenges arise [72].

Coupled with adaptability, it has been proven that a leadership mindset and leadership skills and practices arise from a growth mindset [73]. In this study, 23.81% of the

experts outlined leadership and decision-making as important soft skills needed to work in circularity practices. These are obviously very attractive traits for an employee, particularly in project managers, as stated earlier, as their work will inspire other coworkers around them to increase the quality of their work, encouraging a productive culture in the workplace. Parallel to this, a growth mindset is a facet of resilience [74]. Resilience is a key engineering soft skill in the implementation of CE in smart cities as there will be complex systems being installed that harness the IoT, systems that must then be optimized and maintained, which is not always simple. Issues involving IoT include high energy consumption in IoT devices and poor e-waste management [75], both of which reduce the efficiency of the systems. A resilient individual will continue to overcome setbacks until these problems are mitigated, which, again, will inspire positive organizational behaviors in the work environment [76–80] and will improve the efficiency of this circular system in smart city settings.

6. Conclusions

Technologies and innovations alone cannot promote circular economy adoption in the built environment sector. Similarly, an engineer or a scientist alone cannot stop climate change despite their knowledge of the subject. In addition to technical barriers, there are more hidden challenges to overcome in order to encourage the public to adopt a circular economy, which will reduce the demand for raw resources, greenhouse gas emissions, and, eventually, climate urgency. Engineering skillsets and transferable capabilities span a broad spectrum, covering a multitude of disciplines, ranging from technical engineering expertise to project management and people management, all of which have an impact on project delivery and success in some aspect. This research offers a comprehensive view of the indispensable hard and soft skills that industry experts highlight as essential for their work whilst offering insights into how such skills can be harnessed to tackle the main complex challenges faced, which contribute to low circularity ratings, especially in Europe, and to suggest the steps that can be taken to improve said circularity. Overall, this study indicates skill-based solutions to the key factors that prohibit the development of CE. Our findings are robust, supported by a comprehensive sample size determination based on a review of extensive qualitative data analysis literature, ensuring sufficient data to demonstrate clear trends and effects.

The significance of hard skills in navigating the complex issues that prohibit the development of CE in smart cities is paramount. Our results reveal that background knowledge in engineering and sustainability is the most critical hard skill, according to the experts. Other highly critical hard skills outlined by experts include computing, software and coding, and data analysis and modeling. These skills are crucial regarding digitality in governance, allowing engineers to revise current IoT architectures, with the aim of increasing efficiency. Data analysis and modeling skills are also key in understanding, quantifying, and analyzing BIM software data to act on in order to optimize circular systems. The industry experts also highlight that knowledge in BIM is a valuable hard skill for working in the industry, offering an array of information regarding materials, bills of quantities, energy, emissions, and other LCA information. These benefit CE as they offer an opportunity for circularity analysis involving reviewing buildings at the end of their life, with the aim of renovating or reusing as many materials as possible.

Interestingly, most experts stress that stakeholder engagement is a highly critical soft skill. The correlation between the project manager and stakeholder's trust and the result of the project emphasizes the importance of the right capabilities of the people who implement circularity projects. This is followed by communication skills, which is ranked as highly critical. In addition, the majority of participants state that a second language is preferable.

The ability to life-long learn is often coupled with higher learning capacity, suggesting that the individual will be better able to solve complex problems that will often be encountered in the digitality aspect of CE, such as the optimization and installation of the IoT. It has been proven that great communication skills in education have the potential to influence people's interests and attitudes, which can help to combat one of the biggest issues in implementing CE, namely, educating people on CE practices and behaviors. Management is also ranked as highly critical, as outlined by 52.38% of experts. The related areas of time management, waste management, resource management, and project management all indicate the importance of leadership and interpersonal skills, technical skills, and administrative skills.

With respect to personal development, people with a growth mindset harness the skill of adaptability, being able to attack new and different challenges amidst the ever-changing climate of circularity. Coupled with adaptability, it has been proven that a leadership mindset and leadership skills and practices arise from a growth mindset. A growth mindset indicates resilience, which is a key engineering soft skill in the implementation of CE in smart cities as complex IoT architecture will need to be optimized and maintained, which is not always simple. Issues that reduce the potential of IoT systems include high energy consumption in IoT devices and poor e-waste management; a resilient individual will continue to overcome setbacks until these problems are resolved, which, again, will inspire positive organizational behaviors in the work environment and will improve the efficiency of this circular system in smart city settings.

Author Contributions: Conceptualization, B.N., S.K., D.L., A.D.C., R.A., T.T., and D.B.; methodology, B.N., S.K., D.L., A.D.C., R.A., T.T., and D.B.; software, B.N. and S.K.; validation, B.N. and S.K.; formal analysis, B.N. and S.K.; investigation, B.N., S.K., A.D.C., R.A., T.T., and D.B.; resources, S.K.; data curation, B.N.; writing—original draft preparation, B.N. and S.K.; writing—review and editing, B.N., S.K., A.D.C., R.A., T.T., and D.B.; visualization, B.N.; supervision, S.K.; project administration, S.K.; and funding acquisition, S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This article is based upon work from COST Actions (CircularB—Implementation of Circular Economy in the Built Environment, CA21103; and ECO4ALL - EU Circular Economy Network for All, CA22124), supported by COST (European Cooperation in Science and Technology). The APC is kindly sponsored by the University of Birmingham Library's Open Access Fund.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of the University of Birmingham (approved on 1 December 2023). Ethical review and approval were waived for this study because the participants' information are anonymous and the data were aggregately analyzed, with no linkage to the participants' identities.

Informed Consent Statement: Informed consent was obtained from all participants involved in this study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to the restriction of expert data.

Acknowledgments: The authors wish to gratefully acknowledge all experts, interviewees, and responders, who kindly provided their time and effort for the completion of this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bakıcı, T.; Almirall, E.; Wareham, J. A Smart City Initiative: The Case of Barcelona. *J. Knowl. Econ.* **2013**, *4*, 135–148. [CrossRef]
2. Bajare, D.; Zsembinszki, G.; Stylianos, Y.; Kaewunruen, S.; Cidik, M.S.; Schiller, G.; Zhang, N.; Rizzo, A.; Tambovceva, T.; Hendawy, M.; et al. *Stakeholders' Role, Inter-Relationships, and Obstacles in the Implementation of Circular Economy*; Springer: Berlin/Heidelberg, Germany, 2024; Online publication. [CrossRef]

3. Kaewunruen, S.; Teuffel, P.; Donmez Cavdar, A.; Valta, O.; Tambovceva, T.; Bajare, D. Comparisons of stakeholders' influences, inter-relationships, and obstacles for circular economy implementation on existing building sectors. *Sci. Rep.* **2024**, *14*, 11046. [CrossRef]
4. Kaewunruen, S.; Baniotopoulos, C.; Guo, Y.; Sengsri, P.; Teuffel, P.; Bajare, D. 6D-BIM applications to enrich circular value chains and stakeholder engagement within built environments. In *Proceedings of the 4th International Conference "Coordinating Engineering for Sustainability and Resilience" & Midterm Conference of CircularB "Implementation of Circular Economy in the Built Environment", Timișoara, Romania, 29–31 May 2024*, 1st ed.; Lecture Notes in Civil Engineering; Ungureanu, V., Bragança, L., Baniotopoulos, C., Abdalla, K.M., Eds.; Springer: Berlin/Heidelberg, Germany, 2024; Volume 489, pp. 346–356. [CrossRef]
5. European Commission. Circular Economy Action Plan. 2024. Available online: https://environment.ec.europa.eu/strategy/circular-economy-action-plan_en (accessed on 6 October 2024).
6. Ellen MacArthur Foundation. What Is a Circular Economy? 2024. Available online: https://www.ellenmacarthurfoundation.org/topics/circular-economy-introduction/overview (accessed on 6 October 2024).
7. Eurostat. EU's Circular Material Use Rate Decreased in 2021. 2024. Available online: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20221213-1 (accessed on 6 October 2024).
8. Neale, B. *Civil Engineering Research Project Final Report*; University of Birmingham: Birmingham, UK, 2024.
9. Silva, B.N.; Khan, M.; Han, K. Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities. *Sustain. Cities Soc.* **2018**, *38*, 697–713. [CrossRef]
10. Saheb, T.; Cabanillas, F.J.L.; Higueras, E. The risks and benefits of Internet of Things (IoT) and their influence on smartwatch use. *Span. J. Mark. ESIC* **2022**, *26*, 309–324. [CrossRef]
11. Diogo, P.; Reis, L.P.; Lopes, N.V. Internet of Things: A system's architecture proposal. In Proceedings of the 9th Iberian Conference on Information Systems and Technologies (CISTI), Barcelona, Spain, 18–21 June 2014. [CrossRef]
12. Mohanty, S.P.; Choppali, U.; Kougianos, E. Everything you wanted to know about smart cities: The Internet of things is the backbone. *IEEE Consum. Electron. Mag.* **2016**, *5*, 60–70. [CrossRef]
13. Nam, T.; Pardo, T.A. Conceptualizing smart city with dimensions of technology, people, and institutions. In Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times (dg.o'11), College Park, MD, USA, 12–15 June 2011; Association for Computing Machinery: New York, NY, USA, 2011; pp. 282–291. [CrossRef]
14. UN Press. Only 11 Years Left to Prevent Irreversible Damage from Climate Change, Speakers Warn During General Assembly High-Level Meeting. 2019. Available online: https://press.un.org/en/2019/ga12131.doc.htm (accessed on 11 January 2025).
15. UN Environment Program. Cities and Climate Change. 2024. Available online: https://www.unep.org/explore-topics/resource-efficiency/what-we-do/cities/cities-and-climate-change (accessed on 11 January 2025).
16. Caragliu, A.; Del Bo, C.; Nijkamp, P. Smart Cities in Europe. *J. Urban Technol.* **2011**, *18*, 65–82. [CrossRef]
17. Komninos, N.; Bratsas, C.; Kakderi, C.; Tsarchopoulos, P. Smart City Ontologies: Improving the effectiveness of smart city applications. *J. Smart Cities* **2015**, *1*, 31–46. [CrossRef]
18. Batty, M.; Axhausen, K.W.; Giannotti, F.; Pozdnoukhov, A.; Bazzani, A.; Wachowicz, M.; Ouzounis, G.; Portugali, Y. Smart cities of the future. *Eur. Phys. J. Spec. Top.* **2012**, *214*, 481–518. [CrossRef]
19. Horbach, J.; Rennings, K.; Sommerfeld, K. Circular Economy and Employment, Centre for European Economic Research, Germany. 2015. Available online: https://circulareconomy.europa.eu/platform/en/knowledge/circular-economy-and-employment (accessed on 11 January 2025).
20. Murray, A.; Skene, K.; Haynes, K. The Circular Economy: An Interdisciplinary Exploration of the Concept and Application in a Global Context. *J. Bus. Ethics* **2017**, *140*, 369–380. [CrossRef]
21. Freestone. Three's Not Enough: The 8 Rs of Recycling—Holland Sentinel. 2024. Available online: https://eu.hollandsentinel.com/story/news/environment/2017/11/06/three-x2019-s-not-enough/17281445007/ (accessed on 6 October 2024).
22. Aceleanu, M.I.; Şerban, A.C.; Suciu, M.-C.; Biţoiu, T.I. The Management of Municipal Waste through Circular Economy in the Context of Smart Cities Development. *IEEE Access* **2019**, *7*, 133602–133614. [CrossRef]
23. Circular, B. About Us'. 2024. Available online: https://circularb.eu/about/ (accessed on 6 October 2024).
24. Bai, C.; Ahmadi, H.B.; Moktadir, M.A.; Kusi-Sarpong, S.; Liou, J.J.H. Analyzing the Interactions Among the Challenges to Circular Economy Practices. *IEEE Access* **2021**, *9*, 63199–63212. [CrossRef]
25. Mark Townsend. 'Met Police Vows to Combat Protests Against London's ULEZ Rollout'. 2024. Available online: https://www.theguardian.com/environment/2023/aug/26/met-police-vows-combat-protests-london-ulez-rollout-expansion-low-emission-zone (accessed on 6 October 2024).
26. Ma, L.; Graham, D.J.; Stettler, M.E. Has the ultra low emission zone in London improved air quality? *Environ. Res. Lett.* **2021**, *16*, 124001. [CrossRef]
27. Barros, M.V.; Salvador, R.; Francisco do Prado, G.; Carlos de Francisco, A.; Pielarski, C.M. Circular economy as a driver to sustainable businesses. *Clean. Environ. Syst.* **2020**, *2*, 100006. [CrossRef]
28. Antonetti, P.; Maklan, S. Exploring post consumption guilt and pride in the context of sustainability. *Psychol. Mark.* **2014**, *31*, 717–735. [CrossRef]

29. Webster. £1bn of Government Contracts Given to Tory 'Friends and Donors' During Crisis, The National.Scot. 2020. Available online: https://www.thenational.scot/news/18813681.1bn-government-contracts-given-tory-friends-donors-crisis/ (accessed on 11 January 2025).
30. Grafstrom, J.; Aasma, S. Breaking Circular Economy Barriers. *J. Clean. Prod.* **2021**, *292*, 126002. [CrossRef]
31. Danubianu, M.; Theodorescu, C.; Corneanu, I. Internet of things and the environment—A brief SWOT analysis of IoT/IoTS initiatives. *Present Environ. Sustain. Dev.* **2019**, *13*, 181–190. [CrossRef]
32. Anttiroiko, A.-V. Smart Circular Cities: Governing the Relationality, Spatiality, and Digitality in the Promotion of Circular Economy in an Urban Region. *Sustainability* **2023**, *15*, 12680. [CrossRef]
33. Bu, C.; Shi, D. The emission reduction effect of daily penalty policy on firms. *J. Environ. Manag.* **2021**, *294*, 112922. [CrossRef]
34. Sebastian, I.; Jacobs, B. The Emergence of Relationality in Governance of Climate Change Adaptation. In *The Palgrave Handbook of Climate Resilient Societies*; Brears, R.C., Ed.; Palgrave Macmillan: Cham, Switzerland, 2021. [CrossRef]
35. Hina, M.; Chauhan, C.; Kaur, P.; Kraus, S.; Dhir, A. Drivers and Barriers of Circular Economy business models: Where we are now, and where we are heading. *J. Clean. Prod.* **2021**, *333*, 130049. [CrossRef]
36. Bertino, G.; Kisser, J.; Zeilinger, J.; Langergraber, G.; Fischer, T.; Österreicher, D. Fundamentals of Building Deconstruction as a Circular Economy Strategy for the Reuse of Construction Materials. *Appl. Sci.* **2021**, *11*, 939. [CrossRef]
37. Popova, Y.; Sproge, I. Decision-Making within Smart City: Waste Sorting. *Sustainability* **2021**, *13*, 10586. [CrossRef]
38. Norris, G.A. Social Impacts in Product Life Cycles Towards Life Cycle Attribute Assessment. *Int. J. Life Cycle Assess.* **2006**, *11*, 97–104. [CrossRef]
39. Ochoa Pacheco, P.; Coello-Montecel, D.; Tello, M.; Lasio, V.; Armijos, A. How do project managers' competencies impact project success? A systematic literature review. *PLoS ONE* **2023**, *18*, e0295417. [CrossRef] [PubMed]
40. Rose, K.; Eldridge, S.; Chapin, L. *The Internet of Things: An Overview—Understanding the Issues and Challenges of a More Connected World*; The Internet Society: Reston, VA, USA, 2015.
41. University of Birmingham. The University of Birmingham Partners with Siemens to Create the Smartest University Campus in the World. 2024. Available online: https://intranet.birmingham.ac.uk/student/news/public/summer-2021/the-university-of-birmingham-partners-with-siemens-to-create-the-smartest-university-campus-in-the-world.aspx (accessed on 6 October 2024).
42. Saunders, B.; Sim, J.; Kingstone, T.; Baker, S.; Waterfield, J.; Bartlam, B.; Burroughs, H.; Jinks, C. Saturation in qualitative research: Exploring its conceptualization and operationalization. *Qual. Quant.* **2018**, *52*, 1893–1907. [CrossRef]
43. Delvetool. Practical Guide to Grounded Theory Research. Available online: https://delvetool.com/groundedtheory (accessed on 6 October 2024).
44. Hennig-Thurau, T. Relationship Quality and Customer Retention through Strategic Communication of Customer Skills. *J. Mark. Manag.* **2000**, *16*, 55–79. [CrossRef]
45. Mason, M. Sample Size and Saturation in PhD Studies Using Qualitative Interviews. *Forum Qual. Soc. Res.* **2010**, *11*, 1428. [CrossRef]
46. Hennink, M.M.; Kaiser, B.N.; Marconi, V.C. Code Saturation Versus Meaning Saturation: How Many Interviews Are Enough? *Qual. Health Res.* **2017**, *27*, 591–608. [CrossRef]
47. Marshall, B.; Cardon, P.; Poddar, A.; Fontenot, R. Does Sample Size Matter in Qualitative Research? A Review of Qualitative Interviews in is Research. *J. Comput. Inf. Syst.* **2013**, *54*, 11–22. [CrossRef]
48. Walliman, N. *Your Research Project: Designing, Planning and Getting Started*; Sage Publishing: Thousand Oaks, CA, USA, 2020; 416 pages; ISBN 1526482134.
49. Guest, G.; Bunce, A.; Johnson, L. How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. *Field Methods* **2006**, *18*, 59–82. [CrossRef]
50. Braun, V.; Clarke, V. Using thematic analysis in psychology. *Qual. Res. Psychol.* **2006**, *3*, 77–101. [CrossRef]
51. Holly, F.; Kolar, G.; Berger, M.; Fink, S.; Ogonowski, P.; Schlund, S. Challenges on the way to a circular economy from the perspective of the Austrian manufacturing industry. *Front. Sustain.* **2023**, *4*, 1243374. [CrossRef]
52. University of Cambridge. Languages Still a Major Barrier to Global Science, New Research Finds. 2016. Available online: https://www.cam.ac.uk/research/news/languages-still-a-major-barrier-to-global-science-new-research-finds (accessed on 11 January 2025).
53. Mehdi Mortaheb, M.; Amini, Y.; Younesian, A.H.; Soltani, P. Impacts of Engineering Work Quality on Project Success. *Procedia Soc. Behav. Sci.* **2013**, *74*, 429–437. [CrossRef]
54. Wormley. Engineering Education and the Science and Engineering Workforce—American Society for Engineering Education, Library of Medicine. 2024. Available online: https://www.ncbi.nlm.nih.gov/books/NBK36351/ (accessed on 6 October 2024).
55. Elmezain, M.; Baduruzzaman, W.H.W.; Khoiry, M.A. The Impact of Project Manager's Skills and Age on Project Success. *Braz. J. Oper. Prod. Manag.* **2021**, *18*, 1–16. [CrossRef]
56. Curşeu, P.L.; Semeijn, J.H.; Nikolova, I. Career challenges in smart cities: A sociotechnical systems view on sustainable careers. *Hum. Relat.* **2021**, *74*, 656–677. [CrossRef]

57. Hanes, D.; Grossetete, P.; Salgueiro, G.; Barton, R.; Henry, J.A. *IoT Fundamentals—Networking Technologies, Protocols, and Use Cases for the Internet of Things*; Cisco Press: Indianapolis, IN, USA, 2017; ISBN 9781587144561.
58. Sohime, F.H.; Ramli, R.; Rahim, F.A.; Bakar, A.A. Exploration Study of Skillsets Needed in Cyber Security Field. In Proceedings of the 2020 8th International Conference on Information Technology and Multimedia (ICIMU), Selangor, Malaysia, 24–26 August 2020; pp. 68–72. [CrossRef]
59. Eyeleko, A.H.; Feng, T. A Critical Overview of Industrial Internet of Things Security and Privacy Issues Using a Layer-Based Hacking Scenario. *IEEE Internet Things J.* **2023**, *10*, 21917–21941. [CrossRef]
60. Institution of Civil Engineers. *BIM in Principle and in Practice*; Institution of Civil Engineers: London, UK, 2015.
61. Di Biccari, C.; Abualdenien, J.; Borrmann, A.; Corallo, A. A BIM-Based Framework to Visually Evaluate Circularity and Life Cycle Cost of buildings. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *290*, 012043. [CrossRef]
62. Murtonen, M.; Olkinuora, E.; Tynjälä, P.; Lehtinen, E. "Do I need research skills in working life?": University students' motivation and difficulties in quantitative methods courses. *High Educ.* **2008**, *56*, 599–612. [CrossRef]
63. Lu, S.C.-Y.; Elmaraghy, W.; Schuh, G.; Wilhelm, R. A Scientific Foundation of Collaborative Engineering. *CIRP Ann.* **2007**, *56*, 605–634. [CrossRef]
64. Karlsen, J.T. Forming relationships with stakeholders in engineering projects. *Eur. J. Ind. Eng.* **2008**, *2*, 35–49. [CrossRef]
65. Suthers, D.; Weiner, A. Groupware for developing critical discussion skills. In Proceedings of the CSCL'95: The First International Conference on Computer Support for Collaborative Learning, Bloomington, IN, USA, 17–20 October 1995; pp. 341–348. [CrossRef]
66. Harley, B.; Hart, D. *Language Aptitude and Second Language Proficiency in Classroom Learners of Different Starting Ages*; University of Cambridge: Cambridge, UK, 1997.
67. Bourne, L. Targeted communication: The key to effective stakeholder engagement. *Procedia Soc. Behav. Sci.* **2016**, *226*, 431–438. [CrossRef]
68. Duta, N.; Panisoara, G.; Panisoara, I.-O. The Effective Communication in Teaching. Diagnostic study regarding the academic learning motivation to students. *Procedia Soc. Behav. Sci.* **2015**, *186*, 1007–1012. [CrossRef]
69. Thamhain, H.J. Developing project management skills. *Proj. Manag. J.* **1991**, *22*, 39–44, 53.
70. Crawford, F.A., Jr. Developing Administrative Skills. Available online: https://www.jtcvs.org/article/S0022-5223(00)28250-0/fulltext (accessed on 11 January 2025).
71. Awan, U.; Shamim, S.; Khan, Z.; Ul Zia, N.; Shariq, S.N.; Khan, M.N. Big data analytics capability and decision-making: The role of data-driven insight on circular economy performance. *Technol. Forecast. Soc. Change* **2021**, *168*, 120766. [CrossRef]
72. Burnette, J.L.; Pollack, J.M.; Forsyth, R.B.; Hoyt, C.L.; Babij, A.D.; Thomas, F.N.; Coy, A.E. A Growth Mindset Intervention: Enhancing Students' Entrepreneurial Self-Efficacy and Career Development. *Entrep. Theory Pract.* **2020**, *44*, 878–908. [CrossRef]
73. Kouzes, T.K.; Posner, B.Z. Influence of managers' mindset on leadership behavior. *Leadersh. Organ. Dev. J.* **2019**, *40*, 829–844. [CrossRef]
74. Boullion, A.; Withers, M.; Lippmann, M. Mindsets: Investigating resilience. *Personal. Individ. Differ.* **2021**, *174*, 110669. [CrossRef]
75. Nizetic, S.; Solic, P.; Lopez-de-Ipina, D.; Gonzalez-de-Artaza, D.L.; Patrono, L. Internet of Things (IoT): Opportunities, issues and challenges towards a smart and sustainable future. *J. Clean. Prod.* **2020**, *274*, 122877. [CrossRef]
76. Hartmann, S.; Weiss, M.; Newman, A.; Hoegi, M. Resilience in the Workplace: A Multilevel Review and Synthesis. *Appl. Psychol.* **2019**, *69*, 913–959. [CrossRef]
77. Kaewunruen, S.; Lin, Y.-H.; Rosli, H.; Fan, C.-W.; Pesta, J.; Fohl, F. Digitalisation of Railway Tunnels for Climate Change Adaptation and Enhanced Asset Circularity. *Sustainability* **2024**, *16*, 9708. [CrossRef]
78. Bajare, D.; Zsembinszki, G.; Vaiciukyniene, D.; Kaewunruen, S.; Cidik, M.S.; Tambovceva, T.; Kahraman, I.; Kilic, G.; Cavdar, A.D. CE Management. In *Circular Economy Design and Management in the Built Environment*; Bragança, L., Griffiths, P., Askar, R., Salles, A., Ungureanu, V., Tsikaloudaki, K., Bajare, D., Zsembinszki, G., Cvetkovska, M., Eds.; Springer Tracts in Civil Engineering; Springer: Cham, Switzerland, 2024. [CrossRef]
79. Koohmishi, M.; Kaewunruen, S.; He, X.; Guo, Y. Advancing railway sustainability: Strategic integration of circular economy principles in ballasted track systems. *J. Clean. Prod.* **2025**, *489*, 144713. [CrossRef]
80. Dincă, G.; Milan, A.-A.; Andronic, M.L.; Pasztori, A.-M.; Dincă, D. Does Circular Economy Contribute to Smart Cities' Sustainable Development? *Int. J. Environ. Res. Public Health* **2022**, *19*, 7627. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

MDPI AG
Grosspeteranlage 5
4052 Basel
Switzerland
Tel.: +41 61 683 77 34

MDPI Books Editorial Office
E-mail: books@mdpi.com
www.mdpi.com/books

Disclaimer/Publisher's Note: The title and front matter of this reprint are at the discretion of the Topic Editors. The publisher is not responsible for their content or any associated concerns. The statements, opinions and data contained in all individual articles are solely those of the individual Editors and contributors and not of MDPI. MDPI disclaims responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-5016-7