electronics

Special Issue Reprint

# Advances in Intelligent Data Analysis and Its Applications, 3rd Edition

Edited by
Chao Zhang, Wentao Li, Huiyan Zhang and Tao Zhan

mdpi.com/journal/electronics

MDPI

# Advances in Intelligent Data Analysis and Its Applications, 3rd Edition

# Advances in Intelligent Data Analysis and Its Applications, 3rd Edition

Guest Editors

**Chao Zhang**
**Wentao Li**
**Huiyan Zhang**
**Tao Zhan**

MDPI

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

*Guest Editors*

Chao Zhang
School of Computer and Information Technology
Shanxi University
Taiyuan
China

Wentao Li
College of Artificial Intelligence
Southwest University
Chongqing
China

Huiyan Zhang
National Research Base of Intelligent Manufacturing Service
Chongqing Technology and Business University
Chongqing
China

Tao Zhan
School of Mathematics and Statistics
Southwest University
Chongqing
China

*Editorial Office*
MDPI AG
Grosspeteranlage 5
4052 Basel, Switzerland

This is a reprint of the Special Issue, published open access by the journal *Electronics* (ISSN 2079-9292), freely accessible at: https://www.mdpi.com/journal/electronics/special_issues/9495T2085G.

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

Lastname, A.A.; Lastname, B.B. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

**ISBN 978-3-7258-9373-7 (Hbk)**
**ISBN 978-3-7258-9374-4 (PDF)**
**https://doi.org/10.3390/books978-3-7258-9374-4**

© 2026 by the authors. Articles in this reprint are Open Access and distributed under the Creative Commons Attribution (CC BY) license. The reprint 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

**Chao Zhang, Wentao Li, Huiyan Zhang and Tao Zhan**
Recent Advances in Intelligent Data Analysis and Its Applications, 3rd Edition
Reprinted from: *Electronics* **2026**, *15*, 3594, https://doi.org/10.3390/electronics15163594 . . . . . **1**

**Wenqi Duan, Wenxia Chang, Hui Zhang, Jinchun Peng and Chao Zhang**
Managing Manipulative and Non-Cooperative Behaviors in Group Decision-Making: A Two-Stage Tri-Objective Consensus Optimization Approach
Reprinted from: *Electronics* **2026**, *15*, 3153, https://doi.org/10.3390/electronics15143153 . . . . . **6**

**Heonjong Yoo and Seonggon Choi**
Predicted Longitude and Latitude Information of the Four-Wheel Mobile Platform Using a Gated Recurrent Unit
Reprinted from: *Electronics* **2026**, *15*, 1439, https://doi.org/10.3390/electronics15071439 . . . . . **55**

**Yusong Zhou and Yuanyuan Wang**
Multi-Source Data Fusion for Anime Pilgrimage Recommendation: Integrating Accessibility, Seasonality, and Popularity
Reprinted from: *Electronics* **2026**, *15*, 419, https://doi.org/10.3390/electronics15020419 . . . . . **76**

**Jinbo Qiu, Delong Cui, Zhiping Peng, Qirui Li and Jieguang He**
A Sequence Prediction Algorithm Integrating Knowledge Graph Embedding and Dynamic Evolution Process
Reprinted from: *Electronics* **2025**, *14*, 4922, https://doi.org/10.3390/electronics14244922 . . . . . **98**

**Renjie Fan and Yuanyuan Wang**
AI-Powered Multimodal System for Haiku Appreciation Based on Intelligent Data Analysis: Validation and Cross-Cultural Extension Potential
Reprinted from: *Electronics* **2025**, *14*, 4921, https://doi.org/10.3390/electronics14244921 . . . . . **128**

**Huarong Chen, Yadong Wu, Jing Lei, Weixin Zhao, Guijuan Wang, Jing Liao and Fupan Wang**
Dynamic Topic Analysis and Visual Analytics for Trajectory Data: A Spatial Embedding Approach
Reprinted from: *Electronics* **2025**, *14*, 4873, https://doi.org/10.3390/electronics14244873 . . . . . **164**

**Ping Huang, Huijuan Zhu, Ying Wang, Lili Dai and Lei Zheng**
Enhanced Semantic BERT for Named Entity Recognition in Education
Reprinted from: *Electronics* **2025**, *14*, 3951, https://doi.org/10.3390/electronics14193951 . . . . . **183**

**Yangwu Zhang, Mingxiao Xu and Guohe Li**
A Multi-Level Feature Fusion Network Integrating BERT and TextCNN
Reprinted from: *Electronics* **2025**, *14*, 3508, https://doi.org/10.3390/electronics14173508 . . . . . **201**

**Haoyang Xia and Yuanyuan Wang**
Enhancing Neural Collaborative Filtering for Product Recommendation by Integrating Sales Data and User Satisfaction
Reprinted from: *Electronics* **2025**, *14*, 3165, https://doi.org/10.3390/electronics14163165 . . . . . **220**

**Yanyan Dai and Kidong Lee**
Multi-Sensor Fusion for Autonomous Mobile Robot Docking: Integrating LiDAR, YOLO-Based AprilTag Detection, and Depth-Aided Localization
Reprinted from: *Electronics* **2025**, *14*, 2769, https://doi.org/10.3390/electronics14142769 . . . . . **246**

**Xue Hou, Tingyu Xu and Chao Zhang**
Building Consensus with Enhanced K-means++ Clustering: A Group Consensus Method Based on Minority Opinion Handling and Decision Indicator Set-Guided Opinion Divergence Degrees
Reprinted from: *Electronics* **2025**, *14*, 1638, https://doi.org/10.3390/electronics14081638 . . . . . **264**

**Tymoteusz Miller, Irmina Durlik , Ewelina Kostecka , Polina Kozlovska, Adrianna Łobodzińska, Sylwia Sokołowska and Agnieszka Nowy**
Integrating Artificial Intelligence Agents with the Internet of Things for Enhanced Environmental Monitoring: Applications in Water Quality and Climate Data
Reprinted from: *Electronics* **2025**, *14*, 696, https://doi.org/10.3390/electronics14040696 . . . . . **303**

*electronics*

*Editorial*

# Recent Advances in Intelligent Data Analysis and Its Applications, 3rd Edition

**Chao Zhang [1,*], Wentao Li [2], Huiyan Zhang [3] and Tao Zhan [4]**

[1] School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China
[2] College of Artificial Intelligence, Southwest University, Chongqing 400715, China; drliwentao@gmail.com
[3] National Research Base of Intelligent Manufacturing Service, Chongqing Technology and Business University, Chongqing 400067, China; huiyanzhang@ctbu.edu.cn
[4] School of Mathematics and Statistics, Southwest University, Chongqing 400715, China; zhantao@swu.edu.cn
* Correspondence: czhang@sxu.edu.cn

## 1. Introduction

The accelerated development of artificial intelligence, big data, and the Internet of Things has profoundly expanded the scale, sources, and application scenarios of data, which increasingly exhibit complex characteristics such as high dimensionality, heterogeneity, dynamics, and multimodality. To address the growing demands of data processing, intelligent data analysis has become a core approach for uncovering patterns, extracting knowledge, and supporting decision-making from complex data [1]. Its paradigms are evolving from traditional statistical description and recognition toward integrated intelligent analysis that combines prediction, reasoning, and interaction.

In practical intelligent data analysis, a central challenge lies in the effective management, representation, and fusion of data with diverse structures and semantic levels. Real-world data encompasses not only numerical attributes but also multidimensional information spanning spatio-temporal dynamics, semantics, knowledge, and social behaviors. Knowledge graph embedding represents entities and their relations in compact latent spaces, laying the foundation for relational reasoning and knowledge augmentation [2]. Its integration with large language models facilitates the combination of explicit domain knowledge with the semantic understanding capabilities of generative models [3]. Deep temporal models have advanced the modeling of long-range dependencies and dynamic variations [4], while multimodal large language models offer new pathways for the joint analysis of heterogeneous information [5]. These developments indicate that intelligent data analysis is shifting from single-source modeling toward collaborative multi-source information representation and knowledge-driven analytics.

Researchers have advanced intelligent data analysis along multiple interconnected directions. Knowledge-enhanced learning leverages structured relationships to improve the semantic understanding and interpretability of predictive models. Temporal and trajectory analysis, through modeling spatio-temporal evolution, supports traffic prediction and dynamic topic discovery. Natural language processing and multimodal learning extend analytical capabilities to unstructured scenarios such as legal and educational texts [5]. Multi-granularity intelligent analysis organizes data at different abstraction levels, balancing efficiency, robustness, and interpretability; granular-ball computing, in particular, offers a multi-granularity representation mechanism for robust classification [6]. Recommender systems integrate user behavior and contextual information to deliver personalized services [7]. Social network group decision-making, by analyzing trust and divergence,

enhances the rationality and fairness of consensus processes [8]. The convergence of artificial intelligence and sensor networks strengthens real-time environmental monitoring [9], while multi-sensor fusion leverages complementary information from diverse devices to improve the perceptual and localization reliability of autonomous systems [10].

In the era of big data and intelligence, intelligent data analysis holds significant implications for computer science and socio-economic development. Beyond improving data utilization efficiency, it facilitates practical applications in domains such as judiciary, education, and transportation. However, as intelligent models become increasingly involved in high-stakes decision-making, accuracy alone is no longer sufficient as a performance metric. Computational efficiency, robustness, interpretability, and fairness must also be considered. Explainable artificial intelligence, by revealing the rationale behind model predictions and decisions, enhances transparency and trustworthiness [11]. Therefore, improving analytical performance while strengthening interpretability and reliability has become a critical direction for the sustainable development of intelligent data analysis.

Following the publication of the first two editions, this Special Issue continues the discussion of methodological progress in intelligent data analysis and its practical deployment. The current edition comprises 12 contributions addressing predictive modeling, recommendation, natural language processing, group decision support, multimodal learning, environmental monitoring, and autonomous technologies. Collectively, the papers demonstrate the role of intelligent analytical methods in building systems with improved accuracy, efficiency, interpretability, and adaptability.

## 2. Overview of Contributions

Conventional environmental monitoring is often constrained by intermittent sampling, manual analysis, and limited spatial coverage. Miller et al. (Contribution 1) review the use of artificial intelligence agents within Internet of Things-enabled systems for observing water conditions and climate-related variables. The study summarizes how connected sensors continuously collect environmental variables and how artificial intelligence models support anomaly detection, trend prediction, pollution forecasting, and automated decision-making. It also discusses the complementary roles of reactive, deliberative, and hybrid artificial intelligence agents in environmental applications. Beyond technical benefits, the review examines persistent challenges involving data quality, sensor drift, device interoperability, privacy, cybersecurity, energy supply, and long-term system sustainability. Emerging directions such as edge computing, federated learning, explainable artificial intelligence, blockchain, digital twins, global monitoring networks, and citizen science are also identified as important components of future environmental intelligence.

Large-scale group decision-making requires meaningful subgroup formation and appropriate treatment of minority opinions. Hou et al. (Contribution 2) propose an enhanced K-means++ consensus method that combines Euclidean, Manhattan, and Hamming distances. The framework further incorporates trust propagation, decision indicator sets, and credibility-aware minority opinion management to improve subgroup homogeneity and the transparency of consensus measurement.

Reliable docking is an essential function for autonomous mobile robots deployed in industrial, logistics, and service settings. Vision-only approaches may be affected by illumination variation, motion blur, and partial occlusion, while LiDAR-only systems provide limited semantic information about the docking target. Dai and Lee (Contribution 3) develop a cross-sensor integration framework integrating vision-based tag recognition, depth-supported three-dimensional positioning, and LiDAR-assisted orientation adjustment. A dedicated tag dataset is also assembled to strengthen detection under challenging operating conditions. The detected image coordinates are combined with depth infor-

https://doi.org/10.3390/electronics15163594

mation to estimate the tag position in the robot coordinate system, while symmetrically grouped LiDAR measurements are used to determine orientation deviation. A hybrid PID controller then fuses visual and LiDAR feedback, and a finite-state machine coordinates target search, approach, yaw alignment, and final docking.

E-commerce platforms must provide useful recommendations to both new users without interaction histories and returning users with established behavioral records. Xia and Wang (Contribution 4) propose a dual-strategy framework in which cold-start products are ranked using market demand, duration of sales activity, and customer evaluations, while personalized recommendations are generated through bias neural collaborative filtering. The latter incorporates product-level market and sentiment information into neural user–item interaction modeling.

Legal judgment documents contain both global legal relationships and local factual details that may influence sentencing decisions. Zhang et al. (Contribution 5) introduce a hierarchical textual representation model combining BERT and TextCNN for job-related crime sentencing prediction. BERT captures contextual legal semantics, whereas convolutional kernels extract local features related to specific criminal circumstances and case attributes.

Named entity recognition in educational texts is complicated by specialized terminology, ambiguous entity boundaries, and imbalanced label distributions. Huang et al. (Contribution 6) propose ES-BERT, which integrates BERT-based character representations with TF-IDF-weighted lexical information from an educational domain dictionary. A BiLSTM–CRF architecture and a focal-loss-based objective are further employed to strengthen contextual modeling and sequence labeling.

Large-scale trajectory data contain movement patterns that change across both space and time, whereas conventional topic models often depend on predefined temporal windows and provide limited support for interpreting spatial topic structures. Chen et al. (Contribution 7) develop a dynamic topic analysis and visual analytics framework based on spatial trajectory embeddings. GPS coordinates are first converted into Morton codes so that spatial proximity can be preserved in the token representation. A domain-retrained RoBERTa model is then used to generate contextualized trajectory embeddings, and BERTopic identifies movement topics from these representations. The framework allows temporal windows to be adjusted without retraining the complete topic model. Coordinated visual views, including trajectory maps, topic-word displays, and frequency trends, further support the interpretation of topic composition, spatial distribution, and temporal evolution.

The appreciation of poetry involves linguistic meaning, visual imagery, and emotional atmosphere rather than text alone. Fan and Wang (Contribution 8) present an affect-aware cross-modal framework for interpreting Japanese haiku. The system maps poems into an affective space defined by valence, energy, and dynamism and uses this shared representation to guide image generation and music selection, creating a coordinated textual, visual, and auditory experience.

Sequence prediction models often struggle to capture long-term dependencies while effectively incorporating heterogeneous contextual information. Qiu et al. (Contribution 9) propose the KD4SP framework, which couples graph-based knowledge representations with temporal state evolution. The framework first organizes sequential observations and their temporal relations into a knowledge graph, allowing contextual and semantic associations to be introduced into the prediction process. A feature extraction module then transforms the embedded representations, while the dynamic evolution layer models changes in the sequence over time. Through this combination, the method jointly considers what relationships exist among variables and how these relationships evolve, thereby strengthening prediction for complex sequences over extended horizons.

https://doi.org/10.3390/electronics15163594

Anime pilgrimage recommendation depends not only on user interest but also on practical travel conditions. Zhou and Wang (Contribution 10) develop a multi-source data fusion framework by jointly considering route convenience, seasonal suitability, and popularity signals from multiple platforms. The resulting evidence is evaluated via a multi-attribute ranking scheme to produce situation-sensitive destination rankings that balance cultural appeal with travel feasibility.

Accurate GPS trajectory forecasting is essential for intelligent transportation and mobile services, but recurrent models must balance prediction capability and computational efficiency. Yoo and Choi (Contribution 11) present an enhanced GRU architecture that incorporates speed, heading, and positional attention. These components help the model represent movement direction and turning behavior while emphasizing informative historical positions, making the framework suitable for real-time and resource-constrained applications.

Achieving agreement in social network group decision-making can be severely disrupted by experts who refuse to revise their opinions or strategically exploit influential positions in the network. Duan et al. (Contribution 12) propose DBI-TOCM, a two-stage consensus optimization method with adaptive behavior recognition. The method constructs a more stable trust structure through multi-path decayed trust aggregation that considers both path reliability and structural diversity. It then balances opinion-adjustment expense, agreement level, and fairness in cost allocation within a unified optimization process. In addition, past credibility records and behavioral signals are employed to differentiate temporary resistance, sustained non-cooperation, and strategic manipulation. Differentiated trust decay and cost adjustment mechanisms are subsequently applied, allowing the framework to regulate abnormal behaviors without completely excluding the corresponding experts from the consensus process.

**Author Contributions:** C.Z., W.L., H.Z. and T.Z. worked together in the whole editorial process of the Special Issue, 'Advances in Intelligent Data Analysis and Its Applications, 3rd Edition', published by the journal Electronics. H.Z. and T.Z. drafted this editorial introduction. C.Z. and W.L. reviewed, edited, and finalized the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This editorial was supported in part by the Central Government Guides Local Science and Technology Innovation (No. YDZJSX2024D015); the Open Project Special Fund of Taihang Laboratory in Shanxi Province, China (No. THYF-KFKT-25010500); the Science and Technology Cooperation and Exchange Special Projects of Shanxi (No. 202504041101007); the Sichuan Science and Technology Program (No. 2025ZNSFSC0806); the Top Young Talents of Shanxi "Three Jin" Talents Program (No. SJYC2025334); and the Wenying Young Scholars of Shanxi University.

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

**List of Contributions:**

1. Miller, T.; Durlik, I.; Kostecka, E.; Kozlovska, P.; Łobodzińska, A.; Sokołowska, S.; Nowy, A. Integrating Artificial Intelligence Agents with the Internet of Things for Enhanced Environmental Monitoring: Applications in Water Quality and Climate Data. *Electronics* **2025**, *14*, 696.
2. Hou, X.; Xu, T.; Zhang, C. Building Consensus with Enhanced K-means++ Clustering: A Group Consensus Method Based on Minority Opinion Handling and Decision Indicator Set-Guided Opinion Divergence Degrees. *Electronics* **2025**, *14*, 1638.
3. Dai, Y.; Lee, K. Multi-Sensor Fusion for Autonomous Mobile Robot Docking: Integrating LiDAR, YOLO-Based AprilTag Detection, and Depth-Aided Localization. *Electronics* **2025**, *14*, 2769.
4. Xia, H.; Wang, Y. Enhancing Neural Collaborative Filtering for Product Recommendation by Integrating Sales Data and User Satisfaction. *Electronics* **2025**, *14*, 3165.

https://doi.org/10.3390/electronics15163594

5. Zhang, Y.; Xu, M.; Li, G. A Multi-Level Feature Fusion Network Integrating BERT and TextCNN. *Electronics* **2025**, *14*, 3508.
6. Huang, P.; Zhu, H.; Wang, Y.; Dai, L.; Zheng, L. Enhanced Semantic BERT for Named Entity Recognition in Education. *Electronics* **2025**, *14*, 3951.
7. Chen, H.; Wu, Y.; Lei, J.; Zhao, W.; Wang, G.; Liao, J.; Wang, F. Dynamic Topic Analysis and Visual Analytics for Trajectory Data: A Spatial Embedding Approach. *Electronics* **2025**, *14*, 4873.
8. Fan, R.; Wang, Y. AI-Powered Multimodal System for Haiku Appreciation Based on Intelligent Data Analysis: Validation and Cross-Cultural Extension Potential. *Electronics* **2025**, *14*, 4921.
9. Qiu, J.; Cui, D.; Peng, Z.; Li, Q.; He, J. A Sequence Prediction Algorithm Integrating Knowledge Graph Embedding and Dynamic Evolution Process. *Electronics* **2025**, *14*, 4922.
10. Zhou, Y.; Wang, Y. Multi-Source Data Fusion for Anime Pilgrimage Recommendation: Integrating Accessibility, Seasonality, and Popularity. *Electronics* **2026**, *15*, 419.
11. Yoo, H.; Choi, S. Predicted Longitude and Latitude Information of the Four-Wheel Mobile Platform Using a Gated Recurrent Unit. *Electronics* **2026**, *15*, 1439.
12. Duan, W.; Chang, W.; Zhang, H.; Peng, J.; Zhang, C. Managing Manipulative and Non-Cooperative Behaviors in Group Decision-Making: A Two-Stage Tri-Objective Consensus Optimization Approach. *Electronics* **2026**, *15*, 3153.

## References


1. Sun, X.D.; Zhao, L.X.; Chen, J.Q.; Cai, Y.D.; Wu, D.M.; Huang, J.Z. Non-MapReduce computing for intelligent big data analysis. *Eng. Appl. Artif. Intell.* **2024**, *129*, 107648. [CrossRef]
2. Cao, J.; Fang, J.; Meng, Z.; Liang, S. Knowledge graph embedding: A survey from the perspective of representation spaces. *ACM Comput. Surv.* **2024**, *56*, 159. [CrossRef]
3. Ibrahim, N.; Aboulela, S.; Ibrahim, A.; Kashef, R. A survey on augmenting knowledge graphs (KGs) with large language models (LLMs): Models, evaluation metrics, benchmarks, and challenges. *Discov. Artif. Intell.* **2024**, *4*, 76. [CrossRef]
4. Yang, C.; Wang, Y.; Yang, B.; Chen, J. GRAformer: A gated residual attention transformer for multivariate time series forecasting. *Neurocomputing* **2024**, *581*, 127466. [CrossRef]
5. Liu, H.; Li, C.; Wu, Q.; Lee, Y.J. Visual instruction tuning. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 34892–34916. [CrossRef]
6. Liu, F.; Zhang, Q.; Xia, S.; Xie, Q.; Liao, W.; Zhang, S. A granular-ball generation method based on local density for classification. *Inf. Sci.* **2025**, *717*, 122295. [CrossRef]
7. Hamidi, H.; Moradi, R. Design of a dynamic and robust recommender system based on item context, trust, rating matrix and rating time using social networks analysis. *J. King Saud Univ.-Comput. Inf. Sci.* **2024**, *36*, 101964. [CrossRef]
8. Zhang, C.; Wang, A.; Zhang, J.; Li, W.; Hu, X. T-CT2CRP-SM: The exploration of dockless bike-sharing system rebalancing problems based on multi-modal signals in social networks. *Signal Process.* **2026**, *244*, 110523. [CrossRef]
9. El Khediri, S.; Benfradj, A.; Thaljaoui, A.; Moulahi, T.; Khan, R.U.; Alabdulatif, A.; Lorenz, P. Integration of artificial intelligence (AI) with sensor networks: Trends, challenges, and future directions. *J. King Saud Univ.-Comput. Inf. Sci.* **2024**, *36*, 101892. [CrossRef]
10. Wang, C.; Cao, Z.; Li, J.; Yu, J.; Wang, S. Hierarchical distribution-based tightly-coupled LiDAR inertial odometry. *IEEE Trans. Intell. Veh.* **2024**, *9*, 1423–1435. [CrossRef]
11. Mersha, M.; Lam, K.; Wood, J.; AlShami, A.; Kalita, J. Explainable artificial intelligence: A survey of needs, techniques, applications, and future direction. *Neurocomputing* **2024**, *599*, 128111. [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.

*electronics*

MDPI

*Article*

# Managing Manipulative and Non-Cooperative Behaviors in Group Decision-Making: A Two-Stage Tri-Objective Consensus Optimization Approach

**Wenqi Duan [1], Wenxia Chang [2], Hui Zhang [3], Jinchun Peng [3] and Chao Zhang [2,*]**

[1] Chumin College, Shanxi University, Taiyuan 030006, China; dongwenqi@sxu.edu.cn
[2] School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China; 18536202190@163.com
[3] iGALAXY Information Co., Ltd., Taiyuan 030006, China; zhanghui@igalaxycn.com (H.Z.); pengjc@vip.163.com (J.P.)
* Correspondence: czhang@sxu.edu.cn

**Abstract**

Non-cooperative and manipulative behaviors are strategic actions adopted by experts for diverse reasons, such as rigid preference retention or self-interested exploitation of network structures. To address these challenges, this paper proposes multi-path trust aggregation with decayed trust propagation and a dynamic identification and adaptive penalty mechanism integrating multi-objective optimization to promote the consensus-reaching process in social network group decision-making (SNGDM). First, we propose a multi-path trust aggregation operator with decayed trust propagation, which incorporates a multi-dimensional path analysis framework that evaluates path reliability and structural diversity, thereby building a stable trust network for consensus feedback. Second, a tri-objective consensus optimization model is developed to provide optimal adjustment suggestions by balancing total adjustment cost, consensus degree, and cost fairness. Third, three types of abnormal behaviors are defined, including short-term non-cooperation, persistent non-cooperation, and manipulation. Based on these distinctions, differentiated management mechanisms are introduced. Specifically, a proportional trust decay strategy is applied to non-cooperative behaviors, while an influence-accelerated decay mechanism is designed for manipulative behaviors. In addition, the historical credit index is dynamically linked to unit cost coefficients to force uncooperative and manipulative experts to revise their opinions. Finally, experiments demonstrate the effectiveness of the proposed approach.

**Keywords:** social network group decision-making; q-rung orthopair fuzzy set; trust aggregation; multi-objective optimization; non-cooperative behaviors; manipulative behaviors

## 1. Introduction

Group decision-making (GDM) was introduced to address unstructured and multi-attribute decision problems in which a single decision maker often lacks sufficient knowledge to make an objective choice independently [1,2]. By aggregating evaluations from multiple experts, GDM enables collective decision-making, reduces individual cognitive burden, and mitigates the impact of personal bias [3–5]. However, aggregating individual preferences into a consistent group preference is a fundamental challenge. To formalize this issue, Arrow proposed the impossibility theorem [6], which shows that no aggregation rule can simultaneously satisfy a set of reasonable fairness conditions, thereby revealing

the inherent limitations of perfect preference aggregation and providing a fundamental theoretical benchmark for GDM research [7]. With the development of social media and online platforms, decision-making environments have become increasingly interconnected, and classical GDM frameworks that assume relatively independent experts are no longer sufficient to capture real-world behaviors [8–10]. In practice, experts are embedded in social networks and their opinions are influenced by trust relationships and peer interactions. To address this limitation, social network group decision-making (SNGDM) has been proposed to explicitly model social influence, and trust-based consensus models have been further developed to better characterize opinion dynamics under networked interactions [11,12].

Effective representation of uncertain and hesitant expert preferences is fundamental for SNGDM under multi-granular fuzzy environments. However, cognitive limitations often prevent experts from providing precise numerical evaluations, motivating the development of mathematical tools to model such uncertainty. Early efforts, such as Zadeh's fuzzy set theory [13], only capture vagueness via membership functions but ignore non-membership information. Intuitionistic fuzzy sets [14] address this by introducing non-membership degrees, yet they still struggle to represent high-dimensional or extreme uncertainty. Linguistic models like the 2-tuple [15] and hesitant fuzzy linguistic term sets [16] further support qualitative assessments but remain limited in flexibility when handling complex, non-standard membership relationships. Recent studies have further addressed LGDM in multi-granularity linguistic environments, such as rough integrated asymmetric cloud models for handling heterogeneous linguistic uncertainty and asymmetry [17]. Since the present framework requires a unified preference representation that supports distance measurement, aggregation, and iterative consensus optimization, we adopt $q$-rung orthopair fuzzy sets ($q$-ROFSs) [18,19], which provide a larger decision space and greater flexibility in representing uncertainty by relaxing the constraint on membership and non-membership exponents. This enables more expressive modeling of expert preferences in our framework, making it particularly suitable for the heterogeneous and hesitant evaluations encountered in SNGDM.

Reliable modeling of trust relationships, including both direct and indirect propagation in social networks, is essential for accurately capturing influence structures in SNGDM. In real-world large-scale networks, direct trust relations are often incomplete or sparse, which makes indirect trust propagation through intermediate nodes a necessary component of trust evaluation [20–23]. To address this issue, Wu et al. [24] proposed a t-norm-based trust propagation operator using the Einstein product to aggregate indirect trust along multi-hop paths. However, early propagation models do not explicitly account for the degradation of trust reliability as propagation paths become longer. Subsequent work introduced discount factors [25] and exponential decay mechanisms [26] to model information loss, while more recent data-driven approaches learn decay parameters objectively [27,28]. Despite these advances, most existing methods still rely on linear aggregation assumptions and fail to capture structural heterogeneity in complex networks, particularly the effects of local path vulnerability and redundancy. As a result, the aggregated trust values often deviate from realistic influence structures. Therefore, it is necessary to jointly incorporate path reliability penalization and structural diversity into trust aggregation mechanisms so as to more faithfully reflect the complex propagation patterns in social networks.

Balancing consensus level, adjustment cost, and fairness under bounded rationality is a core challenge in optimization-based consensus-reaching processes in SNGDM. With the diversification of preference expression tools, GDM research shifted from static aggregation strategies to conflict resolution and the consensus-reaching process (CRP). The CRP guides experts through negotiation and feedback, helping them adjust initial opinions to reach

an agreement [29–31]. Because traditional hard consensus requirements cause high costs, Kacprzyk and Fedrizzi [32] established a soft consensus framework allowing for deviations. Focusing on resource allocation and cost minimization, Ben-Arieh and Easton [33] proposed the minimum cost consensus model (MCCM), marking a transition from mechanical feedback loops to optimization-driven dynamics. Subsequent work extended MCCM by incorporating heterogeneous expert weights [34], variable costs reflecting limited willingness to compromise [35], multi-objective fairness-cost trade-offs [36], and robust optimization under uncertainty [37,38]. Despite these advances, existing models typically prioritize one or two objectives while neglecting the inherent triple tension among consensus level, adjustment cost, and fairness. Moreover, most assume fully rational behavior and overlook the bounded rationality that characterizes real-world expert decision-making. Therefore, it is necessary to jointly incorporate consensus level, adjustment cost, fairness, and bounded rationality into the optimization framework of CRP.

Dynamic identification and regulation of non-cooperative and manipulative behaviors are critical for maintaining robustness and stability during consensus evolution in SNGDM. During consensus evolution in social networks, experts often display destructive behaviors that hinder agreement, driven by conflicts of interest or cognitive biases. For instance, some experts may refuse to adjust their opinions due to overconfidence or short-term interests, manifesting as persistent non-cooperative behaviors [39,40]. Others may engage in manipulative behaviors by artificially inflating or depressing alternative scores, exploiting information asymmetry or power imbalances [41,42]. Because these behaviors often remain concealed within complex network topologies, recent studies have developed identification methods using historical performance, cooperation acceptability indices, and multidimensional features [39,41,43]. Once identified, penalizing these behaviors without breaking the underlying trust network becomes essential. To avoid consensus breakdowns caused by static penalties, adaptive strategies have been proposed, including trust propagation transformations, dynamic trust deductions, and Nash bargaining compensation [27,39,43]. Furthermore, recent work has integrated such penalty mechanisms into multi-objective and robust optimization frameworks [36,37]. Despite these advances, existing approaches typically address non-cooperative and manipulative behaviors separately, overlooking their potential interaction and mutual reinforcement. Moreover, most mechanisms assume that experts' behavioral patterns remain static, failing to adapt to the dynamic evolution of trust networks and individual strategies. Therefore, it is necessary to jointly consider the dynamic identification and adaptive regulation of both non-cooperative and manipulative behaviors within a unified consensus framework while accounting for their interdependence and the evolving network context.

Although previous studies have advanced trust mechanisms, consensus evolution, and behavior management, existing frameworks still present three theoretical and methodological limitations in real-world network decision-making environments involving multi-party games:

- Current trust aggregation models lack a complete measure of reliability and topological independence during multi-path propagation. Existing methods ignore the weakest link limit in long propagation paths. They also do not handle overlapping structures and repeating information across multiple trust paths. As a result, trust values are often weakened too much or affected by repeated connections.
- Existing multi-objective optimization models show a clear imbalance between bounded rationality and cost fairness. Traditional models usually assume experts will compromise unconditionally. This ignores the reality that decision-makers prefer to agree with people they trust. At the same time, few existing models jointly incorporate adjustment-cost minimization, post-adjustment consensus maximization,

and Gini-based cost fairness within a bounded-rationality and trust-driven optimization framework.

- Current behavior management methods lack dynamic identification and complete penalty rules for hidden manipulation. Current static methods cannot separate short-term non-cooperation from hidden manipulation, where experts use central network positions to block consensus. In addition, penalty strategies are usually separate from the final consensus optimization process. They lack a flexible control method that links an expert's historical credit, trust flow adjustments, and unit adjustment costs.

To address these problems, this paper proposes a dynamic behavior identification and two-stage tri-objective consensus optimization framework, termed DBI-TOCM, to support consensus reaching in SNGDM under q-ROFS environments. It should be clarified that the novelty of this study does not lie in independently inventing every component used in the framework. Rather, some elements, such as decayed trust propagation, multi-objective consensus optimization, Gini-based fairness measurement, and evolutionary solution procedures, are developed from or inspired by existing studies. The primary contribution of this study is to integrate these elements into a closed-loop consensus optimization framework in which trust propagation, behavior identification, adaptive penalty, and tri-objective consensus optimization are dynamically coupled. Within this framework, several specific methodological extensions are made, as summarized below.

- First, a reliability- and diversity-aware multi-path trust aggregation operator is developed by extending existing decayed trust propagation methods. The single-path decayed propagation value follows the idea of trust decay in prior studies. The mathematical extension of this study lies in the construction of the multi-path aggregation weight, which combines normalized weakest-link reliability and normalized Jaccard-dissimilarity-based structural diversity. Different from conventional decayed propagation models that mainly consider path length or propagated trust strength, the proposed operator jointly considers propagation strength, local path vulnerability, and topological redundancy. This setting limits the effect of fragile or redundant trust paths and provides a more stable trust structure for consensus feedback.
- Second, a two-stage tri-objective consensus optimization model is formulated to balance adjustment cost, post-adjustment consensus degree, and cost fairness under bounded rationality. The novelty does not come from the separate use of cost minimization, consensus maximization, or Gini-based fairness, which have been discussed in previous consensus models. Instead, this study integrates these objectives into a unified q-ROFS-based optimization formulation with a trust-driven reference opinion matrix, adaptive preservation-bound relaxation, and an explicit Pareto solution selection rule. This formulation enlarges the feasible search space when strict preservation bounds lead to infeasibility, while controlling excessive and uneven opinion adjustments.
- Third, a dynamic identification and management mechanism for non-cooperative and manipulative behaviors is developed. A historical credit index is introduced to record experts' behavioral performance across consensus rounds. Compared with prior studies that usually examine non-cooperation or strategic manipulation separately, the proposed mechanism combines the projection-based following degree, historical credit, inter-round opinion change, single-alternative support, and trust-based centrality to distinguish short-term non-cooperation, persistent non-cooperation, and centrality-driven manipulation within the same CRP. In addition, the adaptive penalty strategy links behavior labels and credit indices with trust flow changes and unit adjustment cost coefficients. This mechanism limits the influence of experts identified as non-cooperative or manipulative and supports the subsequent CRP.

https://doi.org/10.3390/electronics15143153

The remainder of this paper is organized as follows. Section 2 introduces the necessary preliminaries. Section 3 presents the proposed multi-path trust aggregation operator with decayed trust propagation. Section 4 details the two-stage tri-objective consensus optimization model. Section 5 describes the identification and management module for non-cooperative and manipulative behaviors. Section 6 presents a numerical case study to validate the feasibility of the proposed model. Section 7 conducts simulation and sensitivity analyses to verify the model's adaptability and stability. Section 8 provides comparative, ablation, and statistical analyses to further demonstrate its effectiveness and superiority. Finally, Section 9 presents the conclusions of this study.

## 2. Preliminaries

In this section, we present the main theoretical foundations of the proposed model. These include q-ROFSs for representing uncertain preferences, social network analysis for modeling interaction structures, and the CRP with the minimum cost consensus framework for describing opinion adjustment and cost-driven consensus evolution in GDM.

### *2.1. q-Rung Orthopair Fuzzy Sets*

Given the inherent complexity of modern decision-making environments and the diverse professional backgrounds of experts, the evaluation information provided for complex alternatives is frequently characterized by hesitation and uncertainty. To effectively manage these uncertain and ambiguous preferences, q-ROFSs can be employed to represent the decision-makers' preference information. Consequently, this section introduces the fundamental concepts of q-ROFSs, including their definition, score function, aggregation operator, and distance measurement [44].

**Definition 1** ([18]). *Let $X$ be a fixed universe of discourse. A q-ROFS $A$ on $X$ is defined as:*

$$A = \{\langle x, \mu_A(x), \nu_A(x)\rangle \mid x \in X\},$$

*where $\mu_A(x)$ and $\nu_A(x)$ denote the membership degree and the non-membership degree of the element $x \in X$ to the set $A$, respectively. These degrees satisfy the conditions $\mu_A(x), \nu_A(x) \in [0,1]$ and $0 \leq (\mu_A(x))^q + (\nu_A(x))^q \leq 1$ for $q \geq 1$. Furthermore, the hesitancy degree of the element $x$ to $A$ is formulated as:*

$$\pi_A(x) = (1 - (\mu_A(x))^q - (\nu_A(x))^q)^{\frac{1}{q}}.$$

For notational convenience, a q-rung orthopair fuzzy number (q-ROFN) is simply denoted as $a = (\mu, \nu)$. The fundamental operational laws for q-ROFNs, such as addition, multiplication, scalar multiplication, and exponentiation, were systematically proposed by Liu and Wang [45].

**Definition 2** ([45]). *Let $a = (\mu, \nu)$ be a q-ROFN. The score function of $a$ is defined as:*

$$Sc(a) = \mu^q - \nu^q. \tag{1}$$

*Simultaneously, the accuracy function of $a$ is defined as:*

$$Ac(a) = \mu^q + \nu^q. \tag{2}$$

To compare the degree of superiority or inferiority among different alternatives, the score and accuracy functions are utilized. Suppose $a_1$ and $a_2$ are any two q-ROFNs; the comparison rules are established as follows:

- If $Sc(a_1) > Sc(a_2)$, then $a_1$ is strictly superior to $a_2$;
- If $Sc(a_1) = Sc(a_2)$, then $a_1$ is superior to $a_2$ if $Ac(a_1) > Ac(a_2)$.

**Definition 3** ([45]). *Let $\beta_k = (\mu_k, \nu_k)$ $(k = 1, 2, \ldots, n)$ be a collection of q-ROFNs, and let $\mathbf{w} = (w_1, w_2, \ldots, w_n)^\top$ be their associated weight vector, satisfying $w_k \in [0, 1]$ and $\sum_{k=1}^{n} w_k = 1$. The q-rung orthopair fuzzy-weighted average (q-ROFWA) operator is given by:*

$$\begin{aligned} q\text{-}ROFWA(\beta_1, \beta_2, \ldots, \beta_n) &= w_1\beta_1 \oplus w_2\beta_2 \oplus \cdots \oplus w_n\beta_n \\ &= \left( \left( 1 - \prod_{k=1}^{n} (1 - (\mu_k)^q)^{w_k} \right)^{\frac{1}{q}}, \prod_{k=1}^{n} (\nu_k)^{w_k} \right). \end{aligned} \tag{3}$$

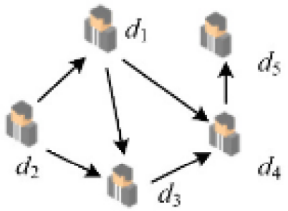
**Definition 4** ([46]). *Consider two q-ROFSs $A = \{A_1, A_2, \ldots, A_n\}$ and $B = \{B_1, B_2, \ldots, B_n\}$, where $A_j = (\mu_{Aj}, \nu_{Aj})$ and $B_j = (\mu_{Bj}, \nu_{Bj})$ are q-ROFNs. The normalized Euclidean distance between A and B is defined as:*

$$D(A, B) = \sqrt{\frac{1}{2n} \sum_{j=1}^{n} \left[ \left( (\mu_{Aj})^q - (\mu_{Bj})^q \right)^2 + \left( (\nu_{Aj})^q - (\nu_{Bj})^q \right)^2 + \left( (\pi_{Aj})^q - (\pi_{Bj})^q \right)^2 \right]}. \tag{4}$$

### 2.2. Social Network Analysis

SNGDM is heavily based on social network analysis (SNA) to model group relationships. A social network consists of experts, their interconnections, and their respective social levels [47]. Let $E = \{e_1, e_2, \ldots, e_K\}$ denote a finite set of experts. The structural relationships among experts can be formalized in three ways: as a graph (nodes representing experts and edges denoting interactions), a sociometric format (an adjacency matrix quantifying relational data), or via algebraic expressions [48]. These three representation formats are summarized in Table 1.

**Table 1.** Three representation formats of social network relationships.

| Graph | Social Matrix | Algebraic |
|---|---|---|
|   | $G = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$ | $e_1Re_3, e_1Re_4$ $e_2Re_1, e_2Re_3$ $e_3Re_4$ $e_4Re_5$ |

Given the inherent transitivity of information in social networks, trust often propagates between individuals lacking direct contact. Based on trust propagation paths, relations among experts are categorized into three distinct types [49]:

- **Direct relationship:** An explicit edge connects two experts, allowing for direct interaction and evaluation;
- **Indirect relationship:** experts are not directly linked but are connected via intermediate mediators, enabling trust to be mathematically derived from existing ties;
- **Irrelevant relationship:** No direct or indirect pathway exists between the experts, indicating an absence of trust or interaction.

### *2.3. Minimum Cost Consensus Model*

In a GDM problem, each expert $e_i$ holds an initial opinion $o_i \in [0,1]$ and is assigned a weight $\omega_i$ (where $\omega_i \geq 0$ and $\sum_{i=1}^{K} \omega_i = 1$). The collective opinion $\bar{o}$ is aggregated via the weighted average: $\bar{o} = \sum_{i=1}^{K} \omega_i o_i$ [34].

Because preference conflicts are inevitable in real-world scenarios, the CRP is applied to help experts negotiate and adjust their opinions. This adjustment naturally incurs a consensus cost. Let $c_i$ represent the unit adjustment cost for expert $e_i$, and $o_i'$ be the adjusted opinion. Recognizing that absolute unanimity is rarely achievable, soft consensus permits an acceptable deviation $\epsilon$ between individual and collective opinions. Building on this, Zhang et al. [34] established the MCCM as follows:

$$\min \sum_{i=1}^{K} c_i |o_i - o_i'|, \tag{5}$$

$$\text{s.t.} \begin{cases} |o_i' - \bar{o}| \leq \epsilon, & i = 1, \ldots, K \\ \bar{o} = \sum_{i=1}^{K} \omega_i \, o_i' \\ \bar{o}, o_i' \in [0,1], & i = 1, \ldots, K \end{cases}$$

The tolerance threshold $\epsilon$ regulates the required consensus level. A smaller $\epsilon$ forces tighter convergence at a higher total cost, whereas a larger $\epsilon$ permits greater opinion dispersion, thereby reducing the necessary adjustments and minimizing costs.

### *2.4. Gini Coefficient*

Introduced by Gini [50] in 1912, the Gini coefficient remains a standard metric for evaluating the fairness of income distribution. It provides an effective means to quantify wealth and income disparities across different regions or industries. The mathematical formulation of the Gini coefficient is:

$$G = \frac{\sum_{k=1}^{N} \sum_{h=1}^{N} |r_k - r_h|}{2N^2 \bar{r}}, \tag{6}$$

where $N$ denotes the total number of industries, $r_k$ and $r_h$ represent the income of the $k$-th and $h$-th industries, respectively, and $\bar{r}$ is the mean income across all industries.

The value of $G$ lies within the interval $[0,1]$, with 0 signifying absolute equality and 1 indicating absolute inequality. Tracking the Gini coefficient helps decision-makers observe distributional shifts following policy implementations, providing an objective basis for designing more equitable economic strategies.

## 3. Multi-Path Trust Aggregation and Expert Weight Calculation

In this section, we first propose a multi-path trust aggregation operator with decayed trust propagation, which incorporates a multi-dimensional path analysis framework that evaluates path reliability and structural diversity. Then, we propose an expert weight calculation method combined with the credit index. For simplicity, the related notations in this paper are listed in Table 2.

**Table 2.** Summary of notation and symbols.

| Symbol | Description |
|---|---|
| $E$ | Set of experts |
| $K$ | Total number of experts |

**Table 2.** *Cont.*

| Symbol | Description |
|---|---|
| $m$ | Number of alternatives |
| $n$ | Number of attributes |
| $t$ | Current consensus round |
| $q$ | Parameter of $q$-ROFS, $q \geq 1$ |
| $\mu, \nu, \pi$ | Membership, non-membership, and hesitancy degrees |
| $D_u^t$ | Opinion matrix of expert $u$ at round $t$ |
| $d_{ij}^{u,t}$ | $q$-ROFN element in matrix $D_u^t$ |
| $D_C^t$ | Collective opinion matrix |
| $AD_u^t$ | Adjusted suggestion matrix for expert $u$ at round $t$ |
| $\tilde{D}_C^t$ | Collective opinion matrix derived from the adjusted opinion matrix |
| $D_{Ru}^t$ | Reference opinion matrix of expert $u$ at round $t$ |
| $P_h$ | $h$th propagation path between two unconnected experts |
| $H$ | Total number of shortest paths |
| $t_i^{(h)}$ | Direct trust score at the $i$th step in path $P_h$ |
| $s_{uv}^t$ | Final trust degree from expert $u$ to $v$ at round $t$ |
| $w_h(i)$ | Trust decay weight allocation function |
| $RS_h$, $\overline{RS}_h$ | Path reliability score and normalized weight |
| $DS_h$, $\overline{DS}_h$ | Structural diversity score and normalized weight |
| $\Lambda_h$ | Unified normalized weight for path $P_h$ |
| $IC_u^t$ | In-degree centrality of expert $u$ |
| $AIC^t$ | Average in-degree centrality of the network |
| $CD_u^t$ | Individual consensus degree of expert $u$ |
| $CD_C^t$ | Collective consensus degree |
| $\epsilon$ | Consensus threshold |
| $\theta_u^t$ | Opinion preservation coefficient of expert $u$ |
| $\theta_{\min}$ | Initial minimum opinion preservation coefficient |
| $\theta_{\min,u}^t$ | Individual minimum opinion preservation coefficient of expert $u$ |
| $PS^*$ | Pareto solution set obtained by NSGA-II |
| $N_P$ | Number of Pareto solutions in $PS^*$ |
| $\widehat{f}_i^{(p)}$ | Normalized value of the $i$th objective for the $p$th Pareto solution |
| $ID^{(p)}$ | Equal-weight normalized ideal-point distance of the $p$th Pareto solution |
| $\boldsymbol{\theta}^t$ | Final preservation coefficient vector selected from $PS^*$ |
| $\Psi_u^t$ | Unit cost coefficient for expert $u$ |
| $r_u^t$ | Absolute adjustment cost for expert $u$ |
| $\bar{r}^t$ | Mean adjustment cost across all experts |
| $\iota$ | Decay factor for boundary relaxation |
| $PFD_u^t$ | Projection-based following degree for expert $u$ |
| $\xi$ | Small positive constant to prevent division by zero |
| $\Phi_u^t$ | Maximum support degree |
| $C_u^t$ | Historical credit index of expert $u$ |
| $w_u^t$ | Influence weight of expert $u$ |
| $\rho$ | Tolerance threshold for recommendation following degree |
| $\beta$ | Credit threshold distinguishing short-term and persistent non-cooperators |
| $\delta$ | High support threshold for identifying manipulators |
| $\gamma_1, \gamma_2, \gamma_3$ | Trust penalty decay coefficients for different non-compliant behaviors |
| $\eta_s, \eta_h$ | Deduction step coefficients for credit index |
| $\kappa$ | Penalty sensitivity parameter for unit cost |

### *3.1. Multi-Path Trust Aggregation Operator with Decayed Trust Propagation*

To address the limitations of uniform propagation weights, path vulnerability, and structural redundancy in traditional methods, this paper proposes a multi-path trust aggregation operator with decayed trust propagation. We formalize a directed social trust network as $\mathcal{G} = (E, L)$. Here, $E = \{e_1, e_2, \ldots, e_K\}$ defines the finite set of experts,

and $L \subseteq \{(e_u, e_v) : e_u, e_v \in E,\ u \neq v\}$ denotes the set of directed edges representing established direct trust relations. The direct trust score from source expert $e_u$ to target expert $e_v$ is quantified as $s_{uv} \in [0, 1]$.

For any two non-adjacent experts $e_u$ and $e_v$ where $(e_u, e_v) \notin L$, we identify a set of shortest trust paths $\mathcal{P} = \{P_1, P_2, \ldots, P_H\}$ connecting them. Each path $P_h \in \mathcal{P}$ consists of $|P_h|$ experts, structured as:

$$P_h : e_u \to e_2^h \to e_3^h \to \cdots \to e_{|P_h|-1}^h \to e_v.$$

When only a single shortest path exists ($H = 1$), the final indirect trust value equals the value of this path, i.e., $s_{uv} = s_{uv}^{(1)}$. In networks with multiple shortest paths ($H > 1$), we apply an aggregation operator. The indirect trust computation consists of three steps.

First, based on the trust decay concept introduced by Gai et al. [27], the indirect trust value for a specific path $P_h$ is calculated as:

$$s_{uv}^{(h)} = \lambda_w(t_1^{(h)}, t_2^{(h)}, \ldots, t_{|P_h|-1}^{(h)}) = \frac{2\prod_{i=1}^{|P_h|-1} w_h(i) t_i^{(h)}}{\prod_{i=1}^{|P_h|-1}\left(2 - w_h(i) t_i^{(h)}\right) + \prod_{i=1}^{|P_h|-1} w_h(i) t_i^{(h)}}, \tag{7}$$

where $t_i^{(h)} - s_{e_i^{(h)} e_{i+1}^{(h)}}$ represents the direct trust score of the $i$-th edge in path $P_h$. The decay weight $w_h(i)$ reduces the trust strength as the topological distance increases [27]:

$$w_h(i) = 1 - \frac{2(i-1)}{|P_h|(|P_h|-1)}, \quad i = 1, 2, \ldots, |P_h| - 1. \tag{8}$$

This decay function accounts for the decrease in trust over intermediary chains, assigning higher priority to direct connections than to distant inferences.

Second, we propose a multi-path trust aggregation operator that measures path reliability and structural diversity.

(1) The reliability of a propagation path depends on its most fragile connection, following the "weakest link" principle. To penalize paths with low-reliability links, the path reliability score $RS_h$ is defined by isolating the minimum direct trust link:

$$RS_h = \min_{1 \le i \le |P_h|-1} t_i^{(h)}, \tag{9}$$

where $t_i^{(h)} = s_{e_i^{(h)} e_{i+1}^{(h)}}$. This score is normalized across all available paths to derive the relative reliability weight $\overline{RS}_h = \frac{RS_h}{\sum_{k=1}^{H} RS_k}$.

(2) Traditional methods lack measures for structural redundancy. If multiple paths share a large number of intermediate nodes, their propagated trust information becomes highly correlated, artificially amplifying the influence of local network clusters. To quantify path diversity, we define the set of intermediate experts for each path as $N_h = \{e_2^{(h)}, e_3^{(h)}, \ldots, e_{|P_h|-1}^{(h)}\}$. The topological overlap between any two distinct paths $P_h$ and $P_k$ is calculated using the Jaccard dissimilarity:

$$J_{hk} = 1 - \frac{|N_k \cap N_h|}{|N_k \cup N_h|}. \tag{10}$$

The structural diversity score $DS_h$ is calculated as the average dissimilarity relative to all alternative paths:

$$DS_h = \frac{1}{H-1} \sum_{k=1, k \neq h}^{H} J_{hk}. \tag{11}$$

https://doi.org/10.3390/electronics15143153

This yields the normalized diversity weight $\overline{DS}_h = \frac{DS_h}{\sum_{k=1}^{H} DS_k}$. Paths with higher diversity scores provide more independent information for trust evaluation.

Third, to determine the final indirect trust value, we integrate the structural measures of reliability and diversity. This mechanism evaluates the structural robustness and independence of each propagation path. The unified normalized weight $\Lambda_h$ for the path $P_h$ is defined as the arithmetic mean of its relative reliability and diversity scores:

$$\Lambda_h = \frac{1}{2}\overline{RS}_h + \frac{1}{2}\overline{DS}_h. \tag{12}$$

Using these structural path weights, the final indirect trust estimate from expert $e_u$ to expert $e_v$ is calculated as the weighted sum of the individual path values:

$$s_{uv} = \sum_{h=1}^{H} \Lambda_h \cdot s_{uv}^{(h)}. \tag{13}$$

This equal-weight distribution evaluates the topological quality of the paths, preventing any single structural anomaly from excessively altering the consensus results such as a highly fragile trust chain or highly redundant paths, while allowing the base propagated trust value $s_{uv}^{(h)}$ to naturally dictate the scale of the final estimation.

### *3.2. Expert Weight Calculation Method Combined with the Credit Index*

In traditional CRP, the evaluation of an expert's influence is limited to their behavior in the current consensus round. To address this limitation and capture the long-term dynamics of expert interactions, we introduce the concept of the historical credit index ($C_u^t$). By incorporating the credit index, experts' behaviors in all consensus rounds can be comprehensively evaluated and tracked. If an expert exhibits non-cooperative or manipulative behaviors, their credit index is penalized and deducted. Conversely, cooperative experts are rewarded with an increase in their credit index. This strategy serves as an effective mechanism to identify and manage experts with long-term non-cooperative tendencies. Furthermore, we combine this credit index with the expert weight determination process.

To calculate the weights of the experts, we define the in-degree centrality $IC_u^t$ of the expert $e_u$ in round $t$, which reflects their trust-based influence within the network:

$$IC_u^t = \frac{1}{K-1} \sum_{v=1, v\neq u}^{K} s_{vu}^t, \tag{14}$$

where $s_{vu}^t$ represents the propagated trust value from expert $e_v$ to $e_u$ in round $t$, and $K$ denotes the total number of experts in the network.

Using the derived in-degree centrality and the historical credit index, we propose a novel dynamic weighting mechanism. The weight of the expert $e_u$ for the subsequent round, denoted $w_u^t$, is calculated by integrating their normalized in-degree centrality with their current credit index $C_u^t$:

$$w_u^t = \frac{IC_u^t \cdot (\varsigma + C_u^t)}{\sum_{v=1}^{K} IC_v^t \cdot (\varsigma + C_v^t)}, \tag{15}$$

where $\varsigma > 0$ is a small smoothing parameter (e.g., $\varsigma = 0.1$). This mechanism effectively reduces the weights of experts who engage in non-cooperative or manipulative behaviors while providing weight growth incentives for cooperative experts. The parameter $\varsigma$ ensures that even experts whose credit index has dropped to zero retain a minimal weight, preventing them from losing their sense of participation or completely leaving the CRP.

https://doi.org/10.3390/electronics15143153

## 4. Two-Stage Tri-Objective Consensus Optimization Model (TOCM)

In GDM, initial expert aggregations rarely meet the predefined consensus threshold. This makes a systematic opinion modification process necessary. Previous approaches typically assume that experts will uniformly follow a central directive to revise their assessments. However, this assumption ignores the bounded rationality, self-preservation tendencies, and economic constraints found in real-world negotiations [22,51,52]. To formalize the opinion negotiation process, this section constructs a TOCM. The model identifies optimal adjustment strategies that trade off conflicting managerial goals while maintaining the psychological bounds of individual experts.

### *4.1. Consensus Measurement Within q-ROFS Environment*

The optimization model is based on a consensus measurement mechanism in the q-ROFS environment. Let $D_u^t = \left(d_{ij}^{u,t}\right)_{m\times n}$ denote the individual evaluation matrix provided by the expert $u$ ($u = 1, 2, \ldots, K$) in round $t$, where $d_{ij}^{u,t} = \left(\mu_{ij}^{u,t}, \nu_{ij}^{u,t}\right)$ is a q-ROFN. Let $D_C^t = \left(d_{ij}^{C,t}\right)_{m\times n}$ represent the collective opinion matrix aggregated via the q-ROFWA operator. Each aggregated element $d_{ij}^{C,t}$ is computed as:

$$d_{ij}^{C,t} = \left(\left(1 - \prod_{u=1}^{K}\left(1 - \left(\mu_{ij}^{u,t}\right)^q\right)^{w_u^t}\right)^{1/q}, \prod_{u=1}^{K}\left(\nu_{ij}^{u,t}\right)^{w_u^t}\right), \tag{16}$$

where $w_u^t$ represents the weight of expert $u$ in round $t$, satisfying $w_u^t \in [0,1]$ and $\sum_{u=1}^{K} w_u^t = 1$ [44].

To quantify opinion differences, the normalized Euclidean distance $D\left(D_u^t, D_C^t\right)$ between individual and collective opinion matrices is calculated using the following formula.

$$D(D_u^t, D_C^t) = \sqrt{\frac{1}{2mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left[\left((\mu_{ij}^{u,t})^q - (\mu_{ij}^{C,t})^q\right)^2 + \left((\nu_{ij}^{u,t})^q - (\nu_{ij}^{C,t})^q\right)^2 + \left((\pi_{ij}^{u,t})^q - (\pi_{ij}^{C,t})^q\right)^2\right]}. \tag{17}$$

The individual consensus degree ($CD_u^t$) for expert $u$ is defined as the complement of this distance:

$$CD_u^t = 1 - D\left(D_u^t, D_C^t\right). \tag{18}$$

The collective consensus degree $CD_C^t$ is then obtained by aggregating all individual consensus degrees:

$$CD_C^t = \sum_{u=1}^{K} w_u^t \cdot CD_u^t. \tag{19}$$

This value determines whether to trigger the optimization model. If $CD_C^t < \epsilon$ (where $\epsilon$ is the consensus threshold), the CRP enters the opinion dynamics phase.

A key challenge in the feedback mechanism is determining how much each expert should adjust their initial opinion to reach an agreement. To model resistance to opinion modification, the preservation coefficient vector $\boldsymbol{\theta}^t = \left[\theta_1^t, \theta_2^t, \ldots, \theta_K^t\right]^\top$ serves as the primary decision variable. The updated opinion for any expert $u \in \{1, \ldots, K\}$ is defined by a convex combination parameterized by $\theta_u^t$:

$$AD_u^t = \theta_u^t D_u^t \oplus (1 - \theta_u^t) D_{Ru}^t, \quad \forall u \in \{1, \ldots, K\}, \tag{20}$$

where $AD_u^t$ denotes the adjusted suggestion matrix, $D_u^t$ denotes the original opinion matrix, and both $\oplus$ and scalar multiplication are performed entrywise according to the operational rules of q-ROFSs. $D_{Ru}^t$ is the reference opinion matrix, which is calculated as:

$$D_{Ru}^{t} = \sum_{v=1, v\neq u}^{K} \tilde{s}_{uv}^{t} D_{v}^{t}, \tag{21}$$

where $\tilde{s}_{uv}^{t} = \frac{s_{uv}^{t}}{\sum_{k=1,k\neq u}^{K} s_{uk}^{t}}$ while $s_{uv}^{t}$ denotes the trust value from expert $u$ to expert $v$, and $\sum$ in Equation (21) represents the $\oplus$ aggregation under q-ROFS operational rules. Under this social network-driven mechanism, experts adjust their opinions based on experts they trust rather than a generic statistical mean. The preservation coefficient $\theta_{u}^{t} \in \left[\theta_{\min,u}^{t}, 1\right]$ controls this effect. Specifically, $\theta_{u}^{t} = 1$ indicates zero adjustment, while $\theta_{u}^{t} = \theta_{\min,u}^{t}$ represents the maximum allowable concession bound for expert $u$.

### *4.2. Formulation of the Tri-Objective Functions*

Opinion adjustment inherently presents a structural problem with competing objectives. To evaluate the associated trade-offs, the proposed TOCM introduces three objective functions that quantify cost, consensus validity, and distributional equity.

#### 4.2.1. Minimization of Adjustment Cost

In practical GDM applications, modifying an expert's initial professional opinion entails economic costs and psychological resistance, such as time expenditure, resource reallocation, and loss of academic prestige. A viable consensus model should therefore avoid requiring excessive or arbitrary adjustment in opinion. The first objective captures these associated costs. By minimizing the total weighted deviation from the initial opinions, the model ensures that the negotiation process remains efficient and acceptable.

$$\min f_{1}(\boldsymbol{\theta}^{t}) = \sum_{u=1}^{K} \Psi_{u}^{t} \cdot D(D_{u}^{t}, AD_{u}^{t}), \tag{22}$$

where $\Psi_{u}^{t} > 0$ denotes the unit cost coefficient for the expert $u$, and $D(\cdot,\cdot)$ is the normalized Euclidean distance. The inclusion of the personalized unit cost coefficient $\Psi_{u}^{t}$ captures the heterogeneity of the experts, as senior experts or those with strict preferences typically pay a higher cost for making concessions. By minimizing this function, the model restricts excessive opinion shifts, which directly lowers the risk of expert dropout or non-compliance.

#### 4.2.2. Maximization of Consensus Degree

Although the model requires that the final consensus degree exceed the consensus threshold $\epsilon$, reaching this lower bound is often insufficient. A threshold-level agreement remains sensitive to minor post-negotiation deviations. To improve the reliability of the result, the second objective maximizes the overall consensus degree based on the adjusted opinions.

$$\max f_{2}(\boldsymbol{\theta}^{t}) = 1 - \sum_{u=1}^{K} w_{u}^{t} \cdot D(AD_{u}^{t}, \tilde{D}_{C}^{t}), \tag{23}$$

where $\tilde{D}_{C}^{t} = \sum_{u=1}^{K} w_{u}^{t} AD_{u}^{t}$ represents the collective opinion derived from the adjusted suggestion matrices, and $\sum$ denotes the aggregation $\oplus$ under the operational rules of q-ROFS. By maximizing $f_{2}$, the model moves beyond the basic feasibility requirement. It forces the adjusted opinions to converge further, which directly improves the stability of the final group decision.

#### 4.2.3. Maximization of Fairness Degree

Traditional consensus models often force outlier experts to make substantial adjustments to close the opinion gap quickly. To address the asymmetric distribution

of these adjustment costs among experts, the proposed model employs a stabilized Gini-index structure:

$$\max f_3(\boldsymbol{\theta}^t) = 1 - \frac{\sum_{u=1}^{K}\sum_{v=1}^{K}|r_u^t - r_v^t|}{2K^2(\bar{r}^t + \xi)}, \tag{24}$$

where $r_u^t = \Psi_u^t \cdot D(D_u^t, AD_u^t)$ represents the absolute cost borne by the expert $u$ and $\bar{r}^t = \frac{1}{K}\sum_{u=1}^{K} r_u^t$ is the mean cost for all experts. The parameter $\xi$ is a small positive constant (e.g., $10^{-8}$) introduced to prevent division by zero when the required adjustments approach zero. By penalizing results that concentrate the adjustment burden on specific individuals, this objective ensures a more even distribution of adjustments across the group.

### *4.3. Constraints and Complete Model*

Subject to the associated constraints, the complete formulation of the TOCM is defined as follows:

$$\begin{aligned} \min \quad & F(\boldsymbol{\theta}^t) = [f_1(\boldsymbol{\theta}^t), -f_2(\boldsymbol{\theta}^t), -f_3(\boldsymbol{\theta}^t)]^\top \\ \text{s.t.} \quad & \begin{cases} f_1(\boldsymbol{\theta}^t) = \sum_{u=1}^{K} \Psi_u^t \cdot D(D_u^t, AD_u^t) \\ f_2(\boldsymbol{\theta}^t) = 1 - \sum_{u=1}^{K} w_u^t \cdot D(AD_u^t, \tilde{D}_C^t) \\ f_3(\boldsymbol{\theta}^t) = 1 - \frac{\sum_{u=1}^{K}\sum_{v=1}^{K}|r_u^t - r_v^t|}{2K^2(\bar{r}^t + \xi)} \\ AD_u^t = \theta_u^t D_u^t + (1-\theta_u^t) D_{Ru}^t \\ D_{Ru}^t = \sum_{v=1, v\neq u}^{K} \tilde{s}_{uv}^t D_v^t \\ \tilde{s}_{uv}^t = \frac{s_{uv}^t}{\sum_{k=1,k\neq u}^{K} s_{uk}^t} \\ \tilde{D}_C^t = \sum_{u=1}^{K} w_u^t AD_u^t \\ r_u^t = \Psi_u^t \cdot D(D_u^t, AD_u^t) \\ \bar{r}^t = \frac{1}{K}\sum_{u=1}^{K} r_u^t \\ \theta_u^t \in [\theta_{\min}, 1] \\ CD_C^t(\boldsymbol{\theta}^t) \geq \epsilon \end{cases} . \end{aligned} \tag{25}$$

The boundary condition applies a minimum preservation coefficient $\theta_{\min}$ to ensure that experts retain a baseline proportion of their initial opinion. The final constraint dictates that the solution must meet the predefined consensus threshold $\epsilon$.

### *4.4. Adaptive Two-Stage Solving Mechanism*

To satisfy constraints $\theta_u^t \in [\theta_{\min}, 1]$ and $CD_C^t(\boldsymbol{\theta}^t) \geq \epsilon$, the model encounters the problem of the feasible region being empty during the solution process. To address this issue, the model employs a two-stage solving mechanism that combines a boundary relaxation strategy with a standard evolutionary algorithm.

#### 4.4.1. The Detection and Relaxation of the Asymmetric Minimum Preservation Coefficient

The first stage establishes a feasible region via an iterative relaxation process. Initially, the lower bound for all experts is set equally to $\theta_{\min,u}^{t,0} = \theta_{\min}$. A feasibility check evaluates the consensus level at this boundary limit, $\boldsymbol{\theta}_{\min}^{t,l} = [\theta_{\min,1}^{t,l}, \ldots, \theta_{\min,K}^{t,l}]^\top$, where $l$ is the iteration index. If $CD_C^t(\boldsymbol{\theta}_{\min}^{t,l}) \geq \epsilon$, the feasible region is not empty, and Stage I ends.If $CD_C^t(\boldsymbol{\theta}_{\min}^{t,l}) < \epsilon$, the algorithm identifies experts whose consensus levels fall below the threshold. The adjustment set $\Omega_{\text{adj}}^l$ is defined as:

$$\Omega_{\text{adj}}^l = \left\{ u \in \{1, \ldots, K\} \mid CD_u^t(\theta_{\min,u}^{t,l}) < \epsilon \right\}. \tag{26}$$

https://doi.org/10.3390/electronics15143153

The individual consensus degree $CD_u^t(\theta_{\min,u}^{t,l})$ in Equation (26) is dynamically evaluated. In each iteration $l$, the temporary adjusted suggestion matrices $\tilde{AD}_u^{t,l}$ are generated for each expert based on the current individual minimum opinion preservation coefficient vector $\boldsymbol{\theta}_{\min}^{t,l}$. Then the temporary collective opinion matrix $\tilde{D}_C^{t,l}$ is calculated from the temporary adjusted suggestion matrices. The metric $CD_u^t(\theta_{\min,u}^{t,l})$ is then strictly calculated as the distance between $\tilde{AD}_u^{t,l}$ and this new dynamic center $\tilde{D}_C^{t,l}$. Consequently, the algorithm identifies bottleneck experts based on the evolving consensus target, rather than an outdated baseline.

The decay factor $\iota \in (0,1)$ then gradually relaxes the lower bound for these identified experts. The update rule for iteration $(l+1)$ is given by:

$$\theta_{\min,u}^{t,l+1} = \begin{cases} \iota \cdot \theta_{\min,u}^{t,l}, & \text{if } u \in \Omega_{adj}^{l} \\ \theta_{\min,u}^{t,l}, & \text{otherwise} \end{cases}. \tag{27}$$

This process repeats until $CD_C^t(\boldsymbol{\theta}_{\min}^{t,l}) \geq \epsilon$. Upon termination in the final iteration $l_{final}$, the resulting individual minimum opinion preservation coefficients are defined as $\theta_{\min,u}^{t} = \theta_{\min,u}^{t,l_{\text{final}}}$ for each expert $u$. By targeting only the bottleneck experts, the algorithm prevents unnecessary modifications for individuals who already meet the requirement, securing a feasible solution without excessive opinion adjustments.

#### 4.4.2. Pareto Optimization via NSGA-II

Once the feasible region is defined by the individual minimum opinion preservation coefficients $[\theta_{\min,u}^{t}, 1]$, the model applies the non-dominated sorting genetic algorithm II (NSGA-II) to identify non-dominated solutions. The algorithm encodes the preservation coefficients $\boldsymbol{\theta}^t$ as chromosomes and initializes a population by uniform sampling within the updated bounds $[\theta_{\min,u}^{t}, 1]$ for each expert $u$. It then uses simulated binary crossover to generate offspring, followed by a polynomial mutation to introduce local variations [36,53].

To enforce the consensus threshold constraint, any individual $\boldsymbol{\theta}^t$ with $CD_C^t(\boldsymbol{\theta}^t) < \epsilon$ receives a penalty, relegating it to the lowest non-domination rank. After evaluating the population across the three objective functions $(f_1, -f_2, -f_3)$, the algorithm sorts the solutions based on non-domination levels and crowding distance. Over successive generations, this process produces a Pareto set $PS^*$ of feasible negotiation strategies.

The NSGA-II implementation used in this study adopts fixed parameter settings. Specifically, the population size is set to $N_p = 100$, the maximum number of generations is $G_{\max} = 200$, the crossover probability is $p_c = 0.9$, the mutation probability is $p_m = 1/K$, the SBX distribution index is $\eta_c = 15$, and the polynomial mutation distribution index is $\eta_m = 20$. Unless otherwise specified, these settings are kept unchanged in experiments.

Since the Pareto set usually contains multiple non-dominated solutions, an explicit final solution selection rule is required to obtain a unique preservation coefficient vector before generating the adjusted suggestion matrices. Different Pareto solutions represent different trade-offs among the total adjustment cost, consensus degree, and fairness degree. This study adopts an equal-weight normalized ideal-point distance rule as the selection rule.

Let the Pareto set obtained by NSGA-II be denoted as $PS^* = \{\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \ldots, \boldsymbol{\theta}^{(N_P)}\}$, where $N_P$ is the number of non-dominated solutions. For each Pareto solution $\boldsymbol{\theta}^{(p)}$, the corresponding objective values are denoted as $f_1^{(p)}$, $f_2^{(p)}$, and $f_3^{(p)}$, where $f_1$ is a cost-type objective to be minimized, while $f_2$ and $f_3$ are benefit-type objectives to be maximized. To make the three objectives comparable, they are normalized into benefit-type indicators as follows:

$$\hat{f}_1^{(p)} = \frac{f_1^{\max} - f_1^{(p)}}{f_1^{\max} - f_1^{\min} + \xi}, \tag{28}$$

$$\hat{f}_2^{(p)} = \frac{f_2^{(p)} - f_2^{\min}}{f_2^{\max} - f_2^{\min} + \xi}, \tag{29}$$

$$\hat{f}_3^{(p)} = \frac{f_3^{(p)} - f_3^{\min}}{f_3^{\max} - f_3^{\min} + \xi}, \tag{30}$$

where $\xi$ is a small positive constant used to avoid division by zero, and $f_i^{\max}$ and $f_i^{\min}$ denote the maximum and minimum values of the $i$-th objective among all solutions in $PS^*$, respectively.

After normalization, the positive ideal point is defined as $I^+ = (1, 1, 1)$. Since the proposed TOCM aims to balance cost efficiency, consensus improvement, and cost fairness, the three objectives are assigned equal weight. The distance between each Pareto solution and the positive ideal point is calculated as:

$$ID^{(p)} = \sqrt{\frac{1}{3}(1 - \hat{f}_1^{(p)})^2 + \frac{1}{3}(1 - \hat{f}_2^{(p)})^2 + \frac{1}{3}(1 - \hat{f}_3^{(p)})^2}. \tag{31}$$

The final Pareto solution is then selected according to the minimum distance rule:

$$p^* = \arg \min_{p=1,\ldots,N_P} ID^{(p)}, \tag{32}$$

and the final preservation coefficient vector is obtained as:

$$\boldsymbol{\theta}^t = \boldsymbol{\theta}^{(p^*)}. \tag{33}$$

If multiple Pareto solutions have the same minimum distance, the solution with the higher consensus degree is selected first; if a tie still exists, the solution with the higher fairness degree is selected; if the tie remains, the solution with the lower total adjustment cost is selected.

Based on the selected final preservation coefficient vector $\boldsymbol{\theta}^t$ obtained by the equal-weight normalized ideal-point distance rule, the adjusted suggestion matrices can be generated for the current consensus round. Specifically, let $\boldsymbol{\theta}^t = [\theta_1^t, \theta_2^t, \ldots, \theta_K^t]^\top$. Substituting $\theta_u^t$ into Equation (20) produces the adjusted suggestion matrices $AD_u^t$, which act as suggestions for experts. This calculation completes the consensus optimization phase. The adjusted suggestion matrices will then serve as suggestions for experts for the next round of the CRP. Based on this, they will submit new opinion matrices for the next round.

## 5. Dynamic Identification and Adaptive Management Framework

In the iterative CRP, experts respond differently to algorithmic feedback-adjusted suggestion matrices. It is important to note that the dynamic identification module described in this section is activated only when the iteration round is greater than 1 ($t > 1$). As defined in Section 4, in the preceding round $t - 1$, the TOCM provides experts with adjusted suggestion matrices $AD_u^{t-1}$. Deviations from these suggestions in the current round $t$ may reflect a reliance on private expertise, or they may indicate non-cooperative and manipulative behaviors.

If the collective consensus degree has not reached the consensus threshold after the experts submit their current updated opinion matrices $D_u^t$, the dynamic identification module is activated. This module identifies inappropriate behaviors on two dimensions:

https://doi.org/10.3390/electronics15143153

behavioral trajectory tracking and intention inference. The objective is to isolate disruptive individuals without penalizing experts who hold genuine minority perspectives.

### *5.1. Dynamic Identification Model for Non-Cooperative and Manipulative Behaviors*

This subsection first establishes the identification criteria for non-cooperators and manipulators, followed by an algorithm to dynamically classify expert behaviors.

#### 5.1.1. Identification of Non-Cooperators

To identify non-cooperative behavior, we measure an expert's willingness to compromise based on their agreement with the moderator's suggestions. Suppose $AD_u^{t-1}$ represents the adjusted suggestion matrix of expert $u$ in the previous round $t-1$, while $D_u^t$ is the opinion matrix submitted by this expert in the current round $t$. Since the q-ROFS framework lacks a strict definition for standard subtraction, we introduce a projection metric based on the normalized Euclidean distance $D(\cdot,\cdot)$. The previous opinion matrix $D_u^{t-1}$, the adjusted opinion matrix $AD_u^{t-1}$, and the newly updated opinion matrix $D_u^t$ form an opinion triangle in a metric space.

**Definition 5.** *We define the projection-based following degree (PFD), denoted by $PFD_u^t$, to measure the effective magnitude of the expert's modification when projected onto the optimal consensus path. Applying the law of cosines within this metric space, PFD is calculated as:*

$$PFD_u^t = \max\left(0, \min\left(1, \frac{D(D_u^{t-1}, D_u^t)^2 + D(D_u^{t-1}, AD_u^{t-1})^2 - D(D_u^t, AD_u^{t-1})^2}{2 \cdot D(D_u^{t-1}, AD_u^{t-1})^2 + \xi}\right)\right), \quad (34)$$

*where $\xi$ is a small positive constant (e.g., $10^{-8}$) used to avoid division by zero if the expert makes no modifications.*

This equation uses a distance-based projection mechanism to assess behavioral compliance. In practice, the mediator assigns a target consensus position, $AD_u^{t-1}$. Because traditional similarity metrics typically evaluate only the angular direction of an adjustment, they exhibit a limitation: an expert might make a small concession in the correct direction to manipulate their compliance score. With the proposed metric-space projection, PFD measures both the direction and the magnitude of the concession. Specifically, if an expert moves directly toward $AD_u^{t-1}$ but covers only a fraction of the required distance, the quadratic distance relationship implies that $PFD_u^t$ corresponds to this proportion (e.g., covering 5% of the distance results in a compliance score of 0.05). Furthermore, the bounds implemented by the max and min functions ensure that orthogonal or divergent adjustments—where the remaining distance to the goal increases—evaluate to zero, while over-adjustments are capped at 1. As a result, this projection evaluates the true cooperation level.

An expert $u$ is considered non-cooperative in round $t$ if their PFD is below the tolerance threshold ($PFD_u^t \leq \rho$) and their individual consensus degree has not reached the target ($CD_u^t < \epsilon$). To differentiate short-term non-cooperative behavior from persistent non-cooperation, we use the expert's historical credit index $C_u^{t-1}$. A non-cooperative expert with $C_u^{t-1} \geq \beta$ is categorized as a short-term non-cooperator (Label 1). Conversely, if $C_u^{t-1} < \beta$, the expert is categorized as a persistent non-cooperator (Label 2), reflecting their persistently low credit.

#### 5.1.2. Identification of Manipulators

Strategic manipulation presents a structural challenge to GDM. Manipulators attempt to bias the final outcome toward specific alternatives [41]. Therefore, we identify manipulators based on the following four behavioral and structural factors:

(1) Experts with a significant opinion mutation against the recommended direction. Manipulators may hide their true intentions in the early stage of the CRP to maintain or enhance their trust-based influence. Once they obtain sufficient structural influence, they may reveal their original preference by sharply changing their opinion toward a specific alternative. To capture this preference-revealing behavior, we compare the actual opinion mutation between two consecutive rounds with the recommended adjustment distance. Specifically, $D(D_u^t, D_u^{t-1})$ measures the actual change in expert (u)'s opinion from round $(t-1)$ to round $(t)$, while $D(D_u^{t-1}, AD_u^{t-1})$ measures the adjustment distance recommended by the moderator in the previous round. If the actual opinion mutation is not smaller than the recommended adjustment distance, and this movement does not sufficiently follow the recommended direction, the behavior indicates a strategic deviation rather than normal cooperation. Therefore, the first subset is defined as $MS_1 = \{ e_u \mid D(D_u^t, D_u^{t-1}) \geq D(D_u^{t-1}, AD_u^{t-1}),\ PFD_u^t \leq \rho \}$;

(2) Experts exhibiting high support for a single alternative. A large disparity in evaluations across alternatives can indicate bias. We compute the maximum support degree $\Phi_u^t$ of expert $u$ by evaluating their scores across all alternatives:

$$\Phi_u^t = \max_i \left( \frac{1}{n} \sum_{j=1}^{n} \mathrm{Sc}(d_{i,j}^t) \right). \tag{35}$$

Experts who favor a specific outcome beyond a tolerance threshold $\delta$ form the subset $MS_2 = \{ e_u \mid \Phi_u^t > \delta \}$;

(3) Experts with high social trust. Manipulators often hold significant influence within the group's network. We measure this via the in-degree centrality index $IC_u^{t-1}$ established in the previous round, and compare it with the network's previous average $AIC^{t-1}$. Experts with higher-than-average trust form the subset $MS_3 = \{ e_u \mid IC_u^{t-1} \geq AIC^{t-1} \}$;

(4) Experts with a low consensus degree. Despite their influence, manipulators resisting opinion convergence will maintain a low consensus degree. This subset is defined as $MS_4 = \{ e_u \mid CD_u^t < \epsilon \}$.

**Definition 6.** *A manipulator is identified when all four criteria are met. The set of experts classified as manipulators (Label 3) is defined as:*

$$MS = \{ e_u \mid e_u \in MS_1 \cap MS_2 \cap MS_3 \cap MS_4 \}. \tag{36}$$

An expert might exhibit high support due to deep domain expertise, strong evidence, or legitimate minority judgment. Similarly, a highly trusted expert whose opinion differs from that of the group may provide valuable professional information rather than strategic interference. Therefore, the proposed rule does not classify an expert as a manipulator based on any single feature. Only when strong single-alternative support, high trust-based influence, a low consensus degree, and a significant opinion mutation against the recommended direction appear simultaneously is the behavior identified as strategic manipulation rather than technical disagreement or legitimate dissent.

#### 5.1.3. Dynamic Identification Algorithm

Based on the formal definitions above, Algorithm 1 summarizes the dynamic behavior identification procedure.

The distinction between Label 2 and Label 3 lies not only in the degree of non-compliance but also in the behavioral trajectory and underlying intention. A persistent non-cooperative expert (Label 2) is mainly characterized by strict preference retention. Such an expert repeatedly refuses to sufficiently follow the moderator's suggestions, as reflected

by a low PFD and a low historical credit index. However, this behavior is essentially a passive resistance to opinion adjustment. The expert may insist on their original judgment due to bounded rationality, professional caution, or a strong preference for private information, but does not necessarily show a strong tendency to promote a single alternative through network influence.

**Algorithm 1** Dynamic behavior identification

**Require:** Current round $t > 1$; updated opinion matrices $D_u^t$; adjusted opinion matrix $AD_u^{t-1}$; thresholds $\rho, \beta, \epsilon, \delta$; expert credit indices $C_u^{t-1}$.
**Ensure:** Behavior label set $\mathcal{B}^t = \{\text{Label}_u^t\}_{u=1}^K$, where Label $\in$ {0 (Cooperative), 1 (Short-term Non-coop), 2 (Persistent Non-coop), 3 (Manipulator)}.
Initialize all expert labels to 0.
Calculate the group consensus degree $CD_C^t$ and individual consensus degrees $CD_u^t$.
**for** each expert $u \in E$ **do**
  Calculate $PFD_u^t$, $\Phi_u^t$, and retrieve the previous network centrality $IC_u^{t-1}$ along with its average $AIC^{t-1}$.
  **if** $CD_u^t < \epsilon$ **then**
    **if** $IC_u^{t-1} \geq AIC^{t-1}$ **and** $\Phi_u^t > \delta$ **and** $D(D_u^t, D_u^{t-1}) \geq D(D_u^{t-1}, AD_u^{t-1})$ **and** $PFD_u^t \leq \rho$ **then**
      $\text{Label}_u^t \leftarrow 3$
    **else**
      **if** $PFD_u^t \leq \rho$ **and** $C_u^{t-1} \geq \beta$ **then**
        $\text{Label}_u^t \leftarrow 1$
      **else if** $PFD_u^t \leq \rho$ **and** $C_u^{t-1} < \beta$ **then**
        $\text{Label}_u^t \leftarrow 2$
      **else**
        $\text{Label}_u^t \leftarrow 0$
      **end if**
    **end if**
  **else**
    $\text{Label}_u^t \leftarrow 0$
  **end if**
**end for**
**return** $\mathcal{B}^t$

By contrast, a manipulative expert (Label 3) exhibits an active strategic deviation. In addition to low consensus and insufficient following of the recommended direction, the expert simultaneously shows strong support for a specific alternative, high trust-based influence, and a significant opinion mutation against the recommended direction. This means that the expert does not merely preserve an original preference; instead, after occupying an influential position in the trust network, the expert reveals or reinforces a targeted preference in a way that may bias the collective outcome. Therefore, the proposed identification method further separates manipulation from ordinary strict preference retention.

This distinction also helps avoid misclassifying legitimate disagreement as manipulation. If an expert disagrees with the group but continues to move toward the recommended direction, lacks high trust-based centrality, or does not exhibit a sudden preference-revealing mutation toward a single alternative, the expert is not classified as a manipulator. Such cases are treated as normal disagreement or non-cooperative behavior depending on their PFD and historical credit. Consequently, the framework applies a proportional penalty to Label 2, while for Label 3 it enforces a rapid weight-reduction mechanism to isolate individuals who exploit their central positions in the trust network.

https://doi.org/10.3390/electronics15143153

*5.2. Adaptive Management of Non-Cooperative and Manipulative Behaviors*

To maintain the trust network while penalizing non-cooperative and manipulative individuals, we propose an adaptive management mechanism. First, this mechanism updates the historical credit index ($C_u^{t-1}$) of experts to reflect the severity of their behaviors. Based on the defined behavior labels, the credit-index increment $\Delta C_u^t$ provides this feedback:

$$C_u^t = \max\left(0, C_u^{t-1} + \Delta C_u^t\right),$$
$$\Delta C_u^t = \begin{cases} -\eta_s \cdot \left(1 - PFD_u^t\right), & \text{if Label} = 1 \\ -\eta_h \cdot \left(1 - PFD_u^t\right), & \text{if Label} = 2 \\ -\eta_h, & \text{if Label} = 3 \\ +\left(CD_u^t - \epsilon\right), & \text{if Label} = 0 \end{cases}. \tag{37}$$

This update mechanism corresponds to our hierarchical penalty strategy. For short-term and persistent non-cooperators (Labels 1 and 2), the trust deduction is scaled by their degree of non-cooperation, quantified as $(1 - PFD_u^t)$. This proportional penalty accounts for the fact that non-cooperative behavior may stem from bounded rationality or a strong preference for private information. By linking the penalty to the divergence of their updated opinion, the system provides a graduated adjustment. It acknowledges minor concessions, encouraging these experts to move toward group consensus without excessively reducing their credit index.

In contrast, manipulators (Label 3) receive a maximum penalty $\eta_h$, regardless of any partial adjustments they make. Manipulators can disrupt the decision-making network. Since they act out of self-interest and exploit accumulated trust, marginal adjustments toward the recommended direction are often strategic efforts to avoid detection and maintain centrality. A proportional penalty would be insufficient and would allow them to reduce system efficiency. Imposing an unconditional deduction ensures a significant reduction of a manipulator's weight, limiting their ability to alter the outcome and ensuring the effectiveness of the CRP.

Following the update of the historical credit index, we apply a trust flow redistribution model to penalize non-cooperative and manipulative experts within the social network. For standard non-cooperators (Labels 1 and 2), the trust they receive undergoes an exponential decay governed by their remaining credit. However, for manipulators (Label 3), we introduce an accelerated decay. Since manipulators exploit their social trust, their penalty is amplified by their own in-degree centrality ($IC_u^{t-1}$), causing a rapid decrease in their trust network:

$$s_{vu}^t = \begin{cases} s_{vu}^{t-1} \cdot \exp\left(-\gamma_{\text{Label}} \cdot \left(1 - C_u^t\right)\right), & \text{if Label} \in \{1, 2\} \\ s_{vu}^{t-1} \cdot \exp\left(-\gamma_3 \cdot \left(1 - C_u^t\right) \cdot \left(1 + IC_u^t\right)\right), & \text{if Label} = 3 \end{cases}, \tag{38}$$

where $\forall v \neq u$, and $\gamma_3 > \gamma_2 > \gamma_1$ represent strictly increasing penalty coefficients.

The released trust $\Delta s_{vu} = s_{vu}^{t-1} - s_{vu}^t$ is reinvested into the remaining cooperative network proportionally to the existing trust relationships. To avoid reallocating trust to experts who are identified as non-cooperative or manipulative in the current round, the eligible redistribution pool is defined as $\mathcal{R}^t = \{i \in E \mid Label_i^t = 0\}$.

Only experts in $\mathcal{R}^t$ are allowed to receive the redistributed trust. Therefore, the updated trust value is calculated as

$$s_{vi}^t = \min\left\{1,\ s_{vi}^{t-1} + \Delta s_{vu} \cdot \frac{s_{vi}^{t-1}}{\sum_{k \in \mathcal{R}^t \setminus v} s_{vk}^{t-1}}\right\}, \quad \forall i \in \mathcal{R}_{v,u}^t. \tag{39}$$

https://doi.org/10.3390/electronics15143153

The truncation in Equation (39) ensures that each updated pairwise trust value remains within $[0, 1]$. In this study, the trust matrix is regarded as a bounded trust-intensity matrix rather than a row-stochastic trust distribution; therefore, no row-sum normalization condition is imposed after redistribution. The proportional redistribution rule is used to preserve expert's local trust structure among eligible cooperative experts, while any unallocated trust is interpreted as local trust loss caused by the identified non-cooperative or manipulative behavior. For experts outside $\mathcal{R}_{v,u}^t$, no positive redistributed trust is assigned in this step. If $\mathcal{R}_{v,u}^t = \emptyset$ or $\sum_{k \in \mathcal{R}_{v,u}^t} s_{vk}^{t-1} = 0$, the released trust $\Delta s_{vu}$ is not redistributed.

Beyond managing the trust network, this credit index is linked to the consensus optimization model to adjust unit cost coefficients. In optimization-based consensus models, unit cost coefficients determine which opinions are modified to reach agreement. We introduce a mechanism to dynamically adjust the unit cost coefficients of experts based on their behavior:

$$\Psi_u^t = \max\left(\Psi_{\min}, \Psi_u^{t-1} \cdot (C_u^t)^{\kappa}\right), \tag{40}$$

where $\kappa > 0$ is the penalty sensitivity parameter controlling the rate of the cost reduction, and $\Psi_{\min}$ (e.g., $10^{-3}$) is a lower bound ensuring the strict convexity and computational stability of the objective function.

This approach departs from traditional static cost settings. It should be emphasized that $\Psi_u^t$ does not represent the actual psychological, professional, reputational, or economic cost privately borne by expert $u$ when changing an opinion. Instead, it is a penalty-adjusted algorithmic unit cost coefficient that determines the degree of protection assigned to the expert's opinion in the optimization objective. Therefore, a decrease in $\Psi_u^t$ does not imply that a low-credit expert is intrinsically easier or less costly to persuade. Rather, it means that the model deliberately reduces the optimization weight associated with modifying that expert's opinion, thereby assigning it a higher adjustment priority.

By coupling the algorithmic cost coefficient $\Psi_u^t$ directly with the credit index, cooperative experts retain relatively stronger protection for their original opinions. Consequently, the optimization algorithm preserves their original opinions, making them less susceptible to adjustment. In contrast, when an expert exhibits manipulative or non-cooperative behavior, the decrease in historical credit reduces their penalty-adjusted algorithmic unit cost coefficient. Consequently, modifying the opinions of these experts contributes less to the cost objective, and their opinions receive a higher adjustment priority in the optimization process. When minimizing the total model-side cost, the algorithm therefore tends to allocate a larger proportion of the required opinion adjustment to manipulative and non-cooperative experts. This limits the influence weight of non-cooperative and manipulative individuals on the final consensus, driving convergence primarily by modifying the opinions of non-cooperative and manipulative experts.

Integrating these penalized parameters into the optimization model directly changes the consensus dynamics. Specifically, the decreased trust weights $s_{iv}^t$ isolate non-cooperators or manipulators within the reference matrix $D_{Ru}^t$, reducing their influence over peers. At the same time, the decreased unit cost coefficient $\Psi_u^t$ leads the objective function $f_1$ to prioritize modifying their opinions to minimize the total cost of the system. As a result, the optimization model allocates larger adjustments to uncooperative experts in subsequent rounds, promoting a global consensus.

### *5.3. Rationale for Behavior-Identification Threshold Setting*

The thresholds in the behavior-identification module are specified according to the value ranges of the corresponding indicators and their functions in distinguishing normal disagreement, non-cooperation, and manipulation.

https://doi.org/10.3390/electronics15143153

Since $PFD_u^t \in [0,1]$ measures the proportion of expert $e_u$'s movement toward the recommended adjustment matrix, the tolerance threshold $\rho$ represents the minimum acceptable following ratio. If $\rho$ is close to zero, only near-complete refusal can be identified. If $\rho$ is set too high, experts who have made partial adjustments may also be classified as non-cooperative. Therefore, $\rho$ should be selected from a low-to-moderate range so that symbolic adjustment and insufficient adjustment can be distinguished from bounded-rational compromise.

The credit threshold $\beta$ is determined by the credit deduction mechanism in Equation (37). Starting from the initial credit level $C_u^1 = 1$, one severe non-cooperative response with $PFD_u^t \approx 0$ reduces the credit index approximately to $1 - \eta_s$, while two consecutive severe non-cooperative responses reduce it approximately to $1 - 2\eta_s$. Therefore, $\beta$ should satisfy

$$1 - 2\eta_s < \beta < 1 - \eta_s.$$

In this study, $\beta = 1 - 1.5\eta_s$ is adopted. This value lies between the credit level after one severe non-cooperative response and that after two consecutive severe non-cooperative responses. Thus, one-time resistance is classified as short-term non-cooperation, whereas repeated resistance is classified as persistent non-cooperation.

The support threshold $\delta$ is used to describe preference concentration toward a single alternative. A high support value alone is insufficient to indicate manipulation, because it may also reflect professional expertise, evidence-based judgment, or legitimate minority opinion. Therefore, $\delta$ is set in the high-support region and is used only as one component of the manipulator identification rule. In the proposed rule, high support must appear together with low consensus degree, high trust-based centrality, significant opinion mutation, and insufficient PFD. This joint condition helps distinguish strategic manipulation from legitimate strong preference.

Overall, $\rho$ defines the minimum PFD, $\beta$ separates short-term and persistent non-cooperation according to the credit evolution rule, and $\delta$ identifies single-alternative preference concentration only under additional behavioral and structural conditions. These constraints provide a basis for threshold setting and improve the interpretability of the identification module.

### *5.4. Selection Process*

When the collective consensus degree $CD_C^t$ reaches the consensus threshold $\epsilon$ via the iterative CRP, the alternatives are ranked through the selection process using the score function and the final collective opinion matrix $D_C^t = (d_{ij}^{C,t})_{m\times n}$, where $d_{ij}^{C,t} = (\mu_{ij}^{C,t}, \nu_{ij}^{C,t})$. The score function is defined as:

$$Sc(d_{ij}^{C,t}) = (\mu_{ij}^{C,t})^q - (\nu_{ij}^{C,t})^q. \tag{41}$$

The overall score for each alternative $a_i$ is then calculated as

$$Sc(d_i^{C,t}) = \frac{1}{n}\sum_{j=1}^{n} Sc(d_{ij}^{C,t}). \tag{42}$$

Finally, the alternatives are ordered in descending order according to their overall scores $Sc(d_i^{C,t})$ to determine the optimal decision. In summary, Algorithm 2 summarizes the execution steps of this CRP, while Figure 1 shows the framework of the proposed method.

https://doi.org/10.3390/electronics15143153

**Algorithm 2** Dynamic behavior identification and two-stage tri-objective consensus model (DBI-TOCM)

**Require:** Expert set $E = \{e_1, e_2, \ldots, e_K\}$ and initial opinion matrices $\{D_u^1\}_{u=1}^K$; initial incomplete trust network $\mathcal{G}^1 = (E, L^1)$; global thresholds $(\epsilon,\ T_{\max})$; behavior parameters $(\rho, \beta, \delta)$; initial credit $C_u^1 = 1.0$ and unit cost $\Psi_u^1 = \Psi_{\text{init}}$ for all $u \in E$.

**Ensure:** Final agreed collective opinion matrix $D_C^*$; optimal alternative ranking.

Apply Equations (7)–(13) to $\mathcal{G}^1$ to establish the complete initial trust matrix $S^1 = [s_{uv}^1]$.
Set iteration counter $t \leftarrow 1$.
Compute the initial expert weight $w_u^1$ incorporating $IC_u^1$ and $C_u^1$ via Equations (14) and (15).
Compute the initial collective opinion matrix $D_C^1$ via Equation (16).
Compute individual consensus degrees $CD_u^1$ and the collective consensus degree $CD_C^1$ via Equations (17)–(19).
**if** $CD_C^1 \geq \epsilon$ **then**
  **goto** Step 30 (Selection Process).
**else**
  Skip behavior management at the initial round.
  Compute feasible bounds $[\theta_{\min,u}^1, 1]$ via asymmetric boundary relaxation using Equations (26) and (27).
  Execute NSGA-II to solve the tri-objective optimization model $F(\boldsymbol{\theta}^1)$ and obtain Pareto-optimal preservation coefficients $\boldsymbol{\theta}^1$.
  Compute the first-round adjusted advice matrix $AD_u^1$ for each expert via Equation (20).
**end if**
**while** $t < T_{\max}$ **do**
  $t \leftarrow t + 1$
  Experts evaluate adjusted suggestion matrices $AD_u^{t-1}$ and submit updated opinion matrices $D_u^t$.
  Compute the new collective opinion matrix $D_C^t$ via Equation (16).
  Compute updated individual consensus degrees $CD_u^t$ and the collective consensus degree $CD_C^t$ via Equations (17)–(19).
  Apply Algorithm 1 to determine the behavior label.
  Compute the updated historical credit index $C_u^t$ via Equation (37).
  Compute the updated trust matrix $S^t$ via Equations (38) and (39).
  Compute the updated unit cost coefficient $\Psi_u^t$ via Equation (40).
  Compute the updated expert weight $w_u^t$ incorporating $S^t$ and $C_u^t$ via Equations (14) and (15).
  **if** $CD_C^t \geq \epsilon$ **then**
    Go to Step 30 (Selection Process).
  **else**
    Execute TOCM with updated adaptive parameters $(D_u^t, S^t, w_u^t, \Psi_u^t)$ to compute new suggested matrices $AD_u^t$.
  **end if**
**end while**
Compute the overall score $Sc(d_i^{C,t})$ for each alternative via Equations (41) and (42).
**return** Alternatives sorted in descending order of $Sc(d_i^{C,t})$.

https://doi.org/10.3390/electronics15143153

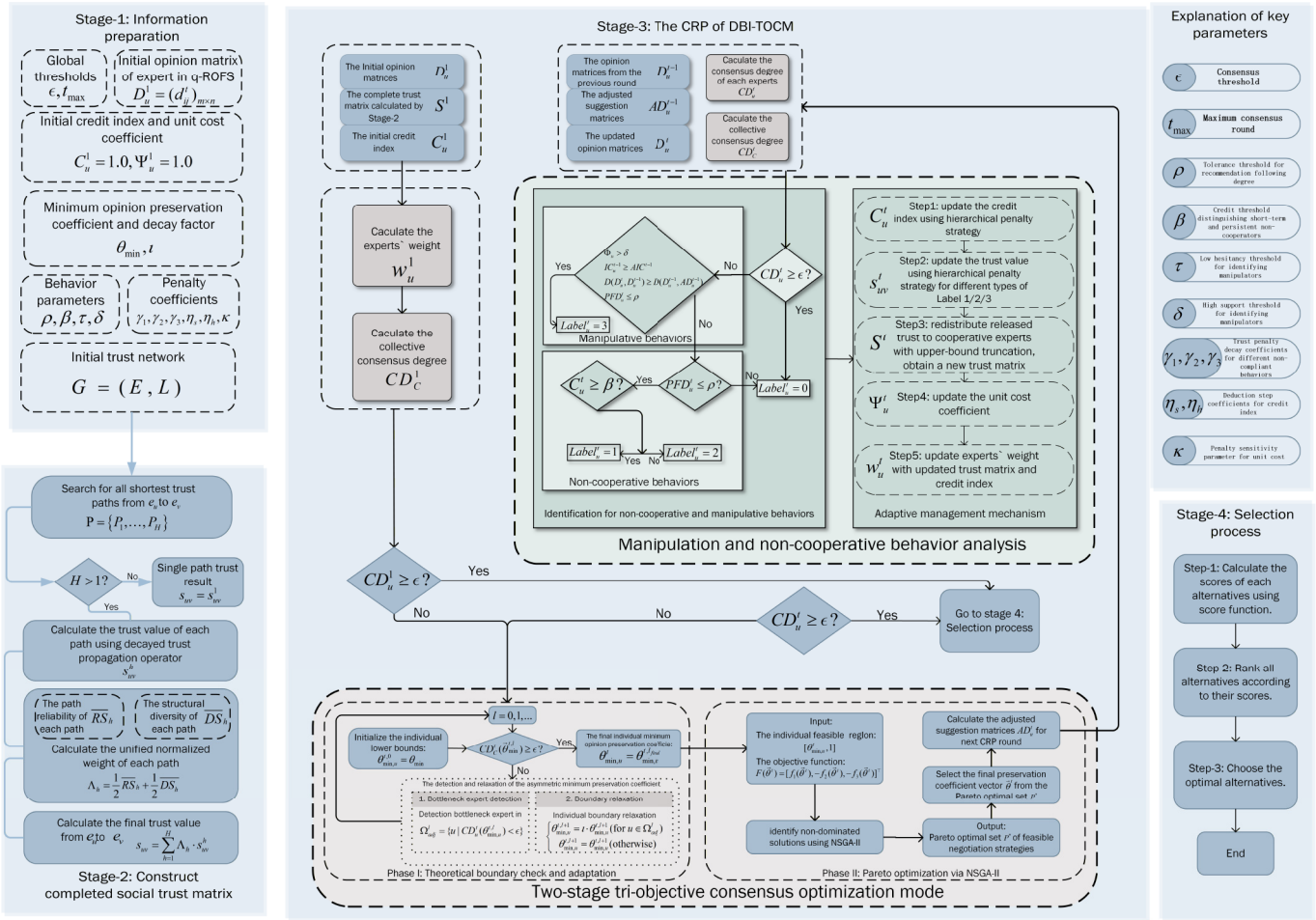



**Figure 1.** The framework of the proposed method.

## 6. Case Study

To validate the feasibility and superiority of the proposed DBI-TOCM, we conduct a simulated case study grounded in a realistic smart city procurement scenario. The initial opinion matrices shown in Appendix A.1 and the incomplete trust network shown in Appendix A.3 are scenario-based synthetic data manually constructed to mirror plausible expert evaluations and organizational relationships in this context. Therefore, these data are used to test the proposed model, rather than to describe an actual procurement case.

The parameter configurations follow the definitions and computational rules of the proposed framework, including the q-ROFS setting, trust propagation mechanism, consensus optimization model, and behavior identification criteria. This design enables the proposed DBI-TOCM to be examined under transparent and reproducible conditions, with particular attention to its dynamic behavior identification and adaptive adjustment mechanisms.

### *6.1. Case Background and Data Construction*

To ensure the scientific and comprehensive selection of the "Smart Medical Resource Dynamic Scheduling Platform," a municipal Health Commission organized a multi-disciplinary decision-making committee. The committee consists of six experts with diverse professional backgrounds. Specifically, $e_1$ is the Director of the Information Center at the Municipal Health Commission; $e_2$ is a provincial public health data expert; $e_3$ serves as a representative from the Budget Department of the Municipal Finance Bureau; $e_4$ is the Director of the Clinical Information Department at a Grade A tertiary hospital; $e_5$ is a

technical consultant for a legacy system maintenance provider; and $e_6$ is the President of a local medical group.

In this context, the GDM process involved multi-attribute assessments and alternative selection. The evaluation is based on three key attributes: technical advancement and stability ($C_1$), deployment and maintenance costs ($C_2$), and interoperability and compatibility of data ($C_3$). Four candidate suppliers entered the final evaluation phase:

- Alternative $A_1$ (Industry Giant): This alternative is developed by a major technology company. It provides an advanced system architecture and high stability. The platform includes many modern functions for data processing. However, the deployment and maintenance expenses are the highest among all options, requiring a large budget from the government.
- Alternative $A_2$ (Local Network): This product is produced by a local enterprise. The system technology is basic and lacks the advanced features found in the other platforms. However, the supplier has established connections with local hospitals and government departments. These existing relationships can facilitate the initial project setup and daily communication.
- Alternative $A_3$ (Compatible Product): This alternative offers standard technical performance. Its core technology is average compared to the other options. The main advantage of this product is its data interoperability and compatibility. It connects easily with the older software systems currently used in various hospitals, achieving seamless data exchange across different medical centers without requiring complete equipment replacements.
- Alternative $A_4$ (Startup Product): This system is developed by a startup company. Its technical capabilities and data interoperability features are basic, providing fewer functions than the other options. The primary advantage of this alternative is its price, offering the lowest cost for both initial deployment and long-term maintenance.

Following the clear definition of these alternatives and attributes, the initial decision information is manually constructed in the form of q-ROFS according to the role characteristics of the six hypothetical experts and the performance features of the four alternatives. Specifically, the initial opinion matrices are designed to reflect the professional concerns of different experts. All q-ROFNs in the initial opinion matrices satisfy the q-ROFS constraint $\mu^q + \nu^q \leq 1$ under the parameter setting used in this case.

Similarly, the initial incomplete trust network is manually assigned based on the assumed historical project collaborations, institutional affiliations, and professional interactions among the six experts. A nonzero off-diagonal entry $s_{uv}$ indicates the direct overt trust degree from expert $e_u$ to expert $e_v$, while a zero entry denotes the absence of a direct overt trust relation rather than distrust. The missing direct trust relations are subsequently inferred through the proposed multi-path trust aggregation operator. Therefore, the initial trust network is intentionally incomplete so as to test the trust propagation component of the proposed method.

In addition, the non-cooperative and manipulative behavioral patterns are embedded by manually designing the subsequent-round opinion matrices. In this case, $e_1$ and $e_3$ are designed to exhibit non-cooperative tendencies in different rounds, while $e_6$ is designed to exhibit manipulative behavior by initially appearing cooperative and then making a sudden preference-revealing opinion shift after gaining sufficient trust-based influence. These behavioral settings are not directly used as model inputs; rather, they are used to generate the updated opinion matrices, and the final behavior labels are identified by the proposed dynamic behavior identification mechanism.

https://doi.org/10.3390/electronics15143153

The initial opinion matrices assigned to each expert are shown in Appendix A.1. Meanwhile, the manually constructed initial incomplete overt trust network $S = [s_{uv}]_{6\times6}$ among experts is shown in Appendix A.3.

### *6.2. Dynamic CRP Under DBI-TOCM*

Before initiating the CRP, the global parameters are configured as follows: $q = 3.0$, $\theta_{\min} = 0.8$, $\iota = 0.9$, $\epsilon = 0.85$, and $T_{\max} = 10$. For the behavior-identification module, the thresholds are calibrated according to the procedure described in Section 5.3. Specifically, the projection tolerance threshold is set as $\rho = 0.3$, which lies in the stable interval identified by the sensitivity analysis. The credit threshold is set as $\beta = 0.70$, derived from $\beta = 1 - 1.5\eta_s$ with $\eta_s = 0.2$, so that one severe deviation is treated as short-term non-cooperation while repeated deviations are classified as persistent non-cooperation. The support threshold is set as $\delta = 0.80$, representing a high single-alternative support level. The trust penalty coefficients are set as $\gamma_1 = 0.5$, $\gamma_2 = 1.5$, and $\gamma_3 = 2.5$, satisfying $\gamma_3 > \gamma_2 > \gamma_1$ to reflect the increasing severity from short-term non-cooperation to manipulation. The credit deduction steps are $\eta_s = 0.2$ and $\eta_h = 0.5$, and the unit cost sensitivity is $\kappa = 1.0$. At the beginning of the CRP, the historical credit index $C_u^1$ and the unit adjustment cost $\Psi_u^1$ for all experts $e_u$ ($u \in \{1, \ldots, 6\}$) are initialized to 1.0.

Based on the initial opinions and the initial trust network, the model executed the following iterative negotiation process:

(1) Round One.

The complete initial trust matrix $S^1$ among experts is constructed based on Equations (7)–(13). Subsequently, the expert weights $w^1$, the collective opinion matrix $D_C^1$, and the collective consensus degree $CD_C^1$ are calculated using Equations (14)–(19). The collective consensus degree is calculated as $CD_C^1 = 0.7999$. Since $CD_C^1 < \epsilon$ and this is the first iteration, behavior identification and management is bypassed. The process directly enters the TOCM to generate the first-round adjusted suggestion matrices $AD_u^1$ for each expert.

(2) Round Two.

Based on the adjusted suggestion matrices $AD_u^1$ provided in Round One, experts submit their second-round opinion matrices $D_u^2$. The new collective consensus degree is calculated as $CD_C^2 = 0.8056$ using the first-round expert weights $w^1$. The dynamic behavior identification and management module is activated. Upon examination, it is found that $e_1$ and $e_3$ exhibit short-term non-cooperative behavior in this consensus round. Based on this, the updated trust matrix $S^2$, historical credit index $C_u^2$, unit cost coefficient $\Psi_u^2$, and expert weights $w_u^2$ are calculated using Equations (37)–(40) and (14) and (15). Specifically, the credit indices of $e_1$ and $e_3$ decrease to $C_1^2 = 0.8000$ and $C_3^2 = 0.8000$, respectively. Since $CD_C^2 < \epsilon$, the TOCM is activated again with the updated parameters to generate the second-round adjusted suggestion matrices $AD_u^2$.

(3) Round Three.

Based on the adjusted suggestion matrices $AD_u^2$ provided in Round Two, experts submit their third-round opinion matrices $D_u^3$. The collective consensus degree is calculated as $CD_C^3 = 0.8164$ using the second-round expert weights $w^2$. The dynamic behavior identification and management module is activated. Upon examination, it is found that $e_3$ exhibits short-term non-cooperative behavior in this consensus round. Based on this, the updated trust matrix $S^3$, historical credit index $C_u^3$, unit cost coefficient $\Psi_u^3$, and expert weights $w_u^3$ are calculated. Specifically, the credit index of $e_3$ decreases to $C_3^3 = 0.6000$. Since $CD_C^3 < \epsilon$, the TOCM is activated with these updated parameters to generate the third-round adjusted suggestion matrices $AD_u^3$.

https://doi.org/10.3390/electronics15143153

(4) Round Four.

Based on the adjusted suggestion matrices $AD_u^3$ provided in Round Three, experts submit their fourth-round opinion matrices $D_u^4$. The collective consensus degree is recalculated as $CD_C^4 = 0.6844$ using the third-round expert weights $w^3$. The dynamic behavior identification and management module is activated. Upon examination, it is found that $e_3$ exhibits persistent non-cooperative behavior and $e_6$ exhibits manipulative behavior in this consensus round. Based on this, the updated trust matrix $S^4$, historical credit index $C_u^4$, unit cost coefficient $\Psi_u^4$, and expert weights $w_u^4$ are calculated. Specifically, the credit index of $e_3$ decreases to $C_3^4 = 0.1000$, while that of $e_6$ decreases to $C_6^4 = 0.5421$. Meanwhile, the unit cost coefficient of $e_3$ decreases to $\Psi_3^4 = 0.0480$. Since $CD_C^4 < \epsilon$, the TOCM is activated again with these updated parameters to generate the fourth-round adjusted suggestion matrices $AD_u^4$.

(5) Round Five.

Based on the adjusted suggestion matrices $AD_u^4$ provided in Round Four, experts submit their fifth-round opinion matrices $D_u^5$. The collective consensus degree is calculated as $CD_C^5 = 0.7842$ using the fourth-round expert weights $w^4$. The dynamic behavior identification and management module is activated. Upon examination, it is found that $e_3$ remains a persistent non-cooperative expert in this consensus round. Based on this, the updated trust matrix $S^5$, historical credit index $C_u^5$, unit cost coefficient $\Psi_u^5$, and expert weights $w_u^5$ are calculated. Specifically, the credit index of $e_3$ decreases to $C_3^5 = 0.0000$, and its unit cost coefficient reaches the lower bound. Since $CD_C^5 < \epsilon$, the TOCM is activated again with these updated parameters to generate the fifth-round adjusted suggestion matrices $AD_u^5$.

(6) Round Six.

Based on the adjusted suggestion matrices $AD_u^5$ provided in Round Five, experts submit their sixth-round opinion matrices $D_u^6$. The collective consensus degree is recalculated as $CD_C^6 = 0.7571$ using the fifth-round expert weights $w^5$. The dynamic behavior identification and management module is activated. Upon examination, it is found that $e_3$ and $e_6$ exhibit persistent non-cooperative behavior in this consensus round. Based on this, the updated trust matrix $S^6$, historical credit index $C_u^6$, unit cost coefficient $\Psi_u^6$, and expert weights $w_u^6$ are calculated. Specifically, both $C_3^6$ and $C_6^6$ decrease to 0.0000, and their unit cost coefficients reach the lower bound. Since $CD_C^6 < \epsilon$, the TOCM is activated again with these updated parameters to generate the sixth-round adjusted suggestion matrices $AD_u^6$.

(7) Round Seven.

Based on the adjusted suggestion matrices $AD_u^6$ provided in Round Six, experts submit their seventh-round opinion matrices $D_u^7$. The collective consensus degree is calculated as $CD_C^7 = 0.9072$ using the sixth-round expert weights $w^6$. The dynamic behavior identification and management module is activated. Upon examination, it is found that $e_3$ and $e_6$ still exhibit persistent non-cooperative behavior in this consensus round. Based on this, the updated trust matrix $S^7$, historical credit index $C_u^7$, unit cost coefficient $\Psi_u^7$, and expert weights $w_u^7$ are calculated. Since $CD_C^7 > \epsilon$, the CRP terminates, and the selection process begins.

The final collective opinion matrix $D_C^*$ is aggregated using the final expert weights $w^7$. The overall scores $Sc(d_i^{C,*})$ for each alternative are then computed using Equations (41) and (42). According to the final scores, the alternatives are ranked as

$$A_3 \succ A_1 \succ A_4 \succ A_2.$$

https://doi.org/10.3390/electronics15143153

Therefore, $A_3$ is selected as the optimal supplier for the Smart Medical Resource Dynamic Scheduling Platform.

In this iterative CRP, the specific data are shown in Appendices A.4 and A.5. Specifically, Appendix A.4 shows the complete trust matrices for 7 rounds of CRP. Table A1 shows the consensus degree for 7 rounds of CRP. Table A2 shows the weight of experts for 7 rounds of CRP. Table A3 shows the credit index of experts for 7 rounds of CRP. Table A4 shows the unit cost coefficient of experts for 7 rounds of CRP. Table A5 shows the behavior label of experts for 7 rounds of CRP.

## 7. Simulation Analysis and Sensitivity Analysis

In this section, we first verify the adaptability and stability of the proposed CRP method under different decision-making scales through simulation experiments. Then, we conduct a sensitivity analysis to evaluate how key parameters influence the performance of the proposed consensus model, including behavioral identification accuracy, total adjustment cost, and cost fairness. The analysis focuses on the tolerance threshold $\rho$ as well as the bounded rationality parameter $\theta_{\min}$ and consensus threshold $\epsilon$, in order to examine their roles in shaping system stability and consensus efficiency.All computational experiments were implemented in Python 3.12.10 using NumPy 2.4.4, Matplotlib 3.10.8, pymoo 0.6.1.6, pandas 3.0.2, SALib 1.5.2, and Seaborn 0.13.2.

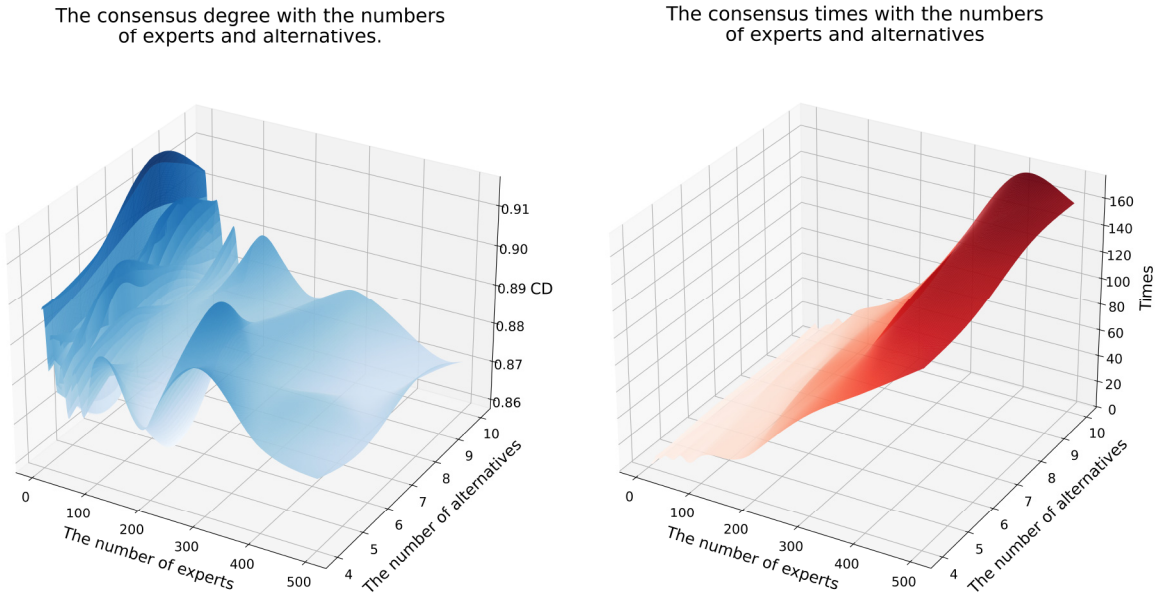
### *7.1. Simulation Analysis*

This section further verifies the adaptability, stability, and scalability of the proposed DBI-TOCM under varying decision-making scales through comprehensive simulation experiments. Specifically, simulated large-scale SNGDM scenarios involving multiple experts and alternative options are constructed. Subsequently, the CRP is executed iteratively according to Algorithm 2 to optimize the initial evaluation matrices of the experts. To evaluate the impact of different group sizes on the consensus level and computational efficiency, the number of experts is configured as $K \in \{5, 10, 20, 30, 50, 75, 100, 150, 200, 300, 400, 500\}$, and the number of alternatives is set as $m \in \{4, 6, 8, 10\}$. For each scenario, the initial q-ROFS evaluation matrices and incomplete social trust networks are randomly generated. Furthermore, to test the robustness of the dynamic identification module, almost 10% of the experts are assigned initial parameters that have non-cooperative or manipulative behavioral tendencies.

As illustrated in Figure 2a, as the number of experts $K$ increases, the model shows a highly stable performance in improving the consensus degree ($CD$). In all scale settings, the final $CD$ strictly satisfies the predefined consensus threshold $\epsilon$ and is significantly higher than the initial $CD$. This verifies that the proposed bounded rational TOCM, coupled with the behavior identification mechanism, possesses a strong consensus-enhancing capability even in massive and highly heterogeneous networks.

In terms of computational efficiency, as shown in Figure 2b, the overall computational time inevitably increases with expansion of $K$ and $m$. This is due to the increased complexity of the multi-path trust aggregation and the population evaluation within the NSGA-II evolutionary process. Nevertheless, the running time remains well within acceptable limits. Overall, the experiment results confirm the robustness of the proposed framework in isolating manipulative behaviors and enhancing consensus while revealing a highly acceptable trade-off between computational cost and group scale in large-scale multi-attribute decision-making contexts.

https://doi.org/10.3390/electronics15143153




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

**Figure 2.** The simulation experiment results of different scales. (**a**) The consensus degree with the numbers of experts and alternatives; (**b**) The consensus times with the numbers of experts and alternatives.

### *7.2. Sensitivity Analysis of the Tolerance Threshold ρ*

In this experiment, we investigate the performance of the proposed behavioral identification mechanism under different tolerance thresholds $\rho$. Specifically, we vary $\rho$ within $[0.05, 0.50]$ to explore its impact on identification accuracy, namely false positives and false negatives, and the total adjustment cost. As illustrated in Figure 3, the total adjustment cost exhibits a characteristic U-shaped trend that aligns closely with the misjudgment distributions. When $\rho$ is small ($\rho \in [0.05, 0.23]$), the system shows excessive tolerance. While no false positives occur, the false negative count remains high, indicating that uncooperative experts can bypass detection through minor, symbolic concessions. This free-riding behavior leaves group disagreements unresolved, forcing genuine cooperators to undertake larger opinion shifts and keeping the total adjustment cost at a high, fluctuating level.

As $\rho$ increases to the interval of $[0.24, 0.33]$, the system achieves an optimal balance. Within this zone, both false positive and false negative counts drop to zero, and the total adjustment cost stabilizes at a global minimum of approximately 14.80. This demonstrates that an appropriate $\rho$ effectively isolates non-cooperators while accommodating the bounded rationality of regular experts. However, as $\rho$ continues to increase beyond 0.33, it is significantly more difficult to achieve a cost-effective consensus. Figure 3 illustrates an increase in false positives alongside a sharp rebound in the total cost. This phenomenon occurs because an excessively strict threshold mechanically penalizes cooperative experts whose reasonable concessions fall short of the extreme limit. Such misjudgments damage the underlying trust network, forcing the optimization model to demand higher global costs to reach an agreement. Therefore, the choice of $\rho$ should be carefully set, with $[0.24, 0.33]$ serving as the optimal range to balance identification accuracy and cost efficiency.

https://doi.org/10.3390/electronics15143153

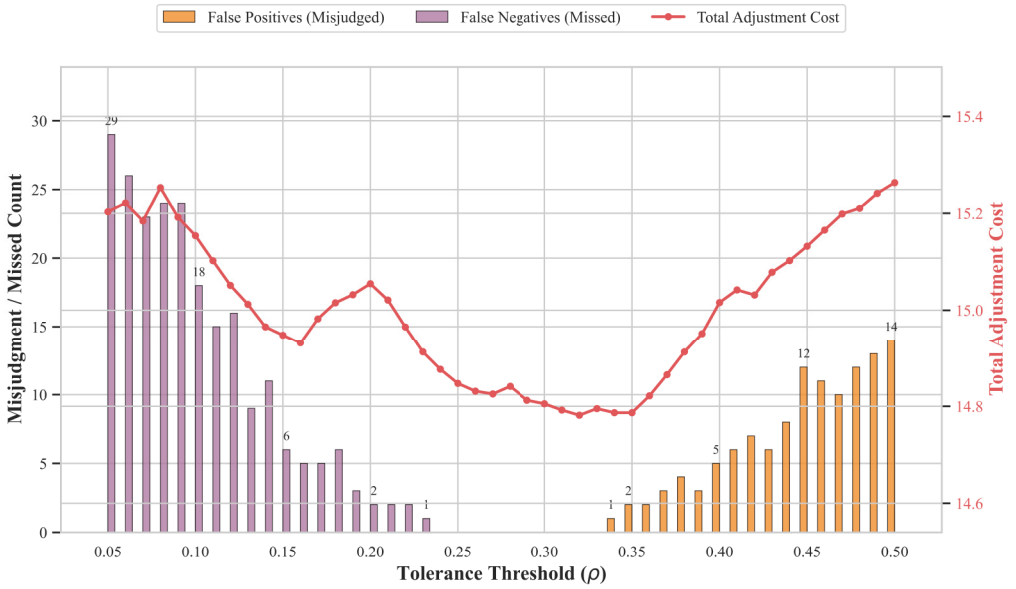



**Figure 3.** Sensitivity analysis of tolerance threshold $\rho$.

### *7.3. Sensitivity Analysis of Bounded Rationality $\theta_{\min}$ and Consensus Threshold $\epsilon$*

In this experiment, we investigate the joint impact of the initial minimum opinion preservation coefficient $\theta_{\min}$ and the consensus threshold $\epsilon$ on the total adjustment cost and cost fairness. Specifically, we vary $\epsilon$ within $[0.80, 0.95]$ and $\theta_{\min}$ within $[0.50, 0.95]$ to explore their effects on the consensus outcomes. Figure 4 presents the outcomes of these variations.

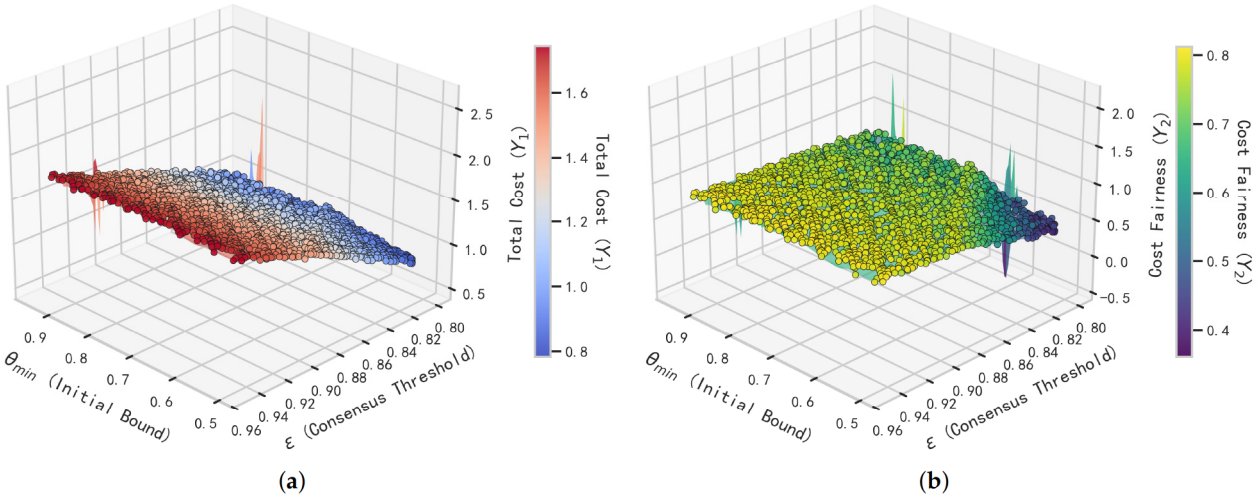



**Figure 4.** Sensitivity analysis of bounded rationality $\theta_{\min}$ and consensus threshold $\epsilon$. (**a**) Total adjustment cost; (**b**) Cost fairness.

As illustrated in Figure 4a, the consensus threshold $\epsilon$ exerts a dominant influence on the total cost. When $\epsilon$ increases from 0.80 to 0.95, the total cost exhibits a strict upward trend. Furthermore, when both $\epsilon$ and $\theta_{\min}$ are at high levels, the total cost surface rises sharply. This indicates that high expert preservation bounds interact strongly with high consensus requirements, leading to a significant increase in the cost required to reach an agreement.

Figure 4b illustrates the corresponding trend of cost fairness under the same parameter variations. In the high-parameter region, the fairness measure drops significantly. This phenomenon occurs because, under extreme conflict, the system compels specific experts to make disproportionate concessions to break the deadlock, resulting in an inequitable cost

https://doi.org/10.3390/electronics15143153

distribution. In practice, this suggests that a decision system cannot simultaneously achieve a high consensus threshold, maintain high individual preservation bounds, and ensure low cost with high fairness. If a high consensus threshold is required, facilitators may need to implement a pre-negotiation mechanism to lower the experts' initial preservation bounds to mitigate the negative impact on cost and fairness.

## 8. Comparative Analysis

To provide a comprehensive analysis and further verify the practicality and effectiveness of the proposed model, this section carries out experiments from multiple perspectives, including comparative analysis, ablation analysis, and statistical analysis.

All baseline methods were implemented under the same initial opinion matrices, initial incomplete trust networks, consensus threshold, maximum iteration number, q-ROFS environment, and stopping criteria. Unless a parameter is method-specific, the same parameter value was used across all methods. The CRP terminates when the collective consensus degree reaches $\epsilon$ or when $T_{\max}$ is reached. In module-level comparisons, only the target module is replaced, while the remaining components are kept unchanged.

### *8.1. Comparative Analysis of Different Trust Propagation Methods*

Yuan et al. [25] proposed a 2-tuple linguistic trust propagation operator with a static discount factor, and aggregated multi-path trust via induced ordered weighted averaging based on knowledge certainty degree. Zhao et al. [54] proposed a 2-tuple linguistic trust propagation operator with continuous exponential decay using the Einstein operator, and aggregated multi-path trust via weighted average based on path complete knowledge degree and path length. Gai et al. [27] proposed a trust propagation operator with objective trust decay using a non-increasing weight allocation function and Einstein product, and aggregated multi-path trust via weighted average based on propagated trust values. Compared with them, we propose a multi-path decayed trust aggregation method incorporating path reliability and structural diversity. The influence of various trust propagation methods on weight allocation and CRP is illustrated in Figure 5 and Table 3, respectively.

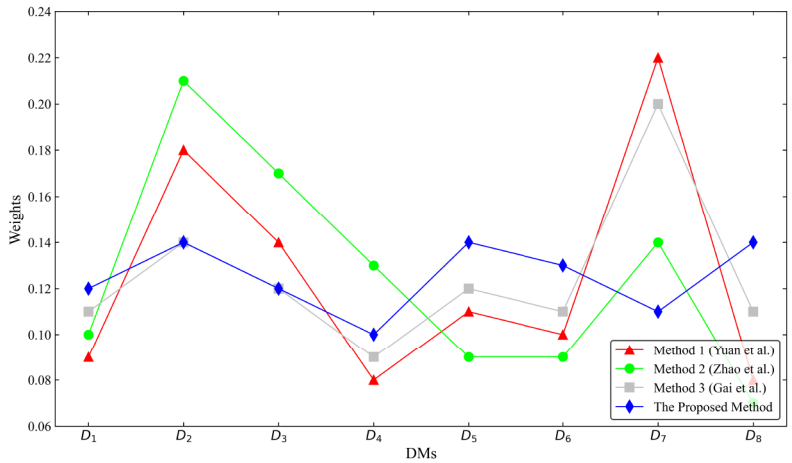



**Figure 5.** Weights of experts based on different trust propagation methods [25,27,54].

**Table 3.** Impact of different trust propagation methods on the CRP.

| Methods | Total Adjustment Cost | Consensus Degree | Fairness Degree |
|---|---|---|---|
| Yuan et al. [25] | 14.5413 | 0.873 | 0.760 |
| Zhao et al. [54] | 16.1577 | 0.894 | 0.784 |
| Gai et al. [27] | 17.3805 | 0.876 | 0.743 |
| Proposed method | 15.3628 | 0.896 | 0.852 |

As shown in Figure 5, different trust propagation methods yield distinct indirect trust levels between experts. This naturally changes how weights are allocated among the experts. Compared to the other approaches, the proposed method calculates significantly different weights for $D_k(k = 2,3,7)$. This difference is driven by two key mechanisms. First, it evaluates structural diversity $DS_h$ to suppress redundant paths. Second, it uses path reliability $RS_h$ to penalize fragile connections. Together, these steps reflect a much more realistic trust propagation process.

Table 3 presents the total adjustment cost, consensus degree and fairness degree of integrating these different trust methods into the TOCM. The proposed method generates a lower total adjustment cost than [27,54]. Meanwhile, it achieves the highest consensus degree and fairness degree. This improvement can be attributed to more reliable trust-based weight and reference opinion construction. By penalizing fragile paths and reducing redundant propagation information, the proposed method avoids overestimating unreliable indirect trust, thereby improving consensus and achieving fairer adjustment allocation.

### *8.2. Comparative Analysis of Different Optimization Models*

Based on the features in Table 4, we compare the models across three dimensions:

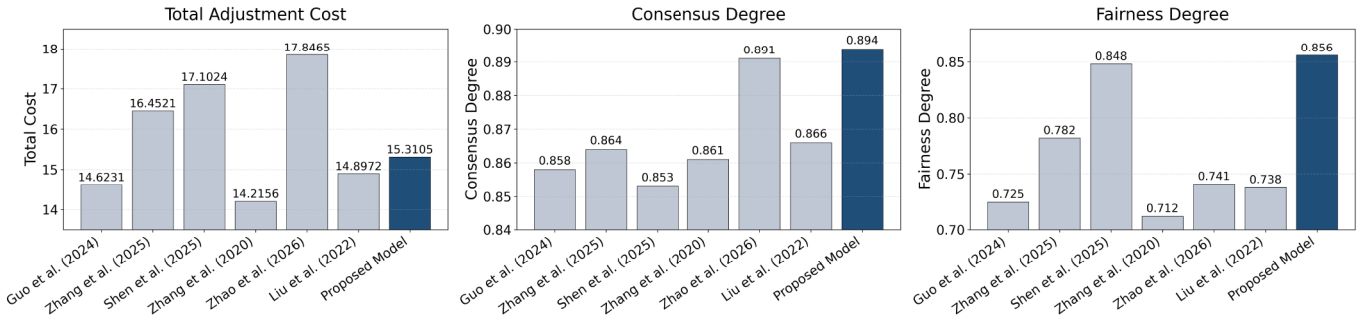
**Table 4.** Comparison of the proposed model with existing consensus optimization models.

| Models | Optimization Objectives | Objective Functions | Unit Adjustment Cost Mechanism | Fairness Constraints |
|---|---|---|---|---|
| Zhang et al. [55] | Single-objective | Minimize total cost/adjustment | Static | None |
| Liu et al. [56] | Single-objective | Minimize total adjustment distance | Static | None |
| Guo et al. [57] | Single-objective | Minimize total consensus cost | Dynamic (Centrality-based) | None |
| Zhang et al. [58] | Bi-level | Upper: Min cost; Lower: Max utility | Static | None (Global) |
| Shen et al. [36] | Bi-objective | Min total cost + Min Gini coefficient | Static | Yes (Gini coefficient) |
| Zhao et al. [54] | Bi-objective | Min total cost + Max consensus degree | Static | None |
| Proposed model | Tri-objective | Min total cost + Max posterior consensus + Max cost fairness | Dynamic (Behavior-driven penalty) | Yes (Gini coefficient) |

1. Traditional models usually focus on single-objective (e.g., Zhang et al. [55]; Liu et al. [56]) or bi-objective optimization (e.g., Zhao et al. [54]). They try to minimize the total cost or balance cost with the consensus degree. This focus on efficiency often causes an uneven cost distribution. Cooperative experts end up bearing most of the adjustment burden. Shen et al. [36] added fairness as a second objective, but they did not optimize the posterior consensus degree. Our model uses a tri-objective framework. It considers cost, consensus level, and fairness at the same time.
2. Many current models use a static unit adjustment cost (e.g., Zhang et al. [58]; Shen et al. [36]; Zhao et al. [54]). With static costs, these models struggle to handle non-cooperative behaviors. They often force normal experts to change their opinions to reach the consensus threshold. Guo et al. [57] used a dynamic mechanism, but it is based on network centrality instead of user behavior. Our model adds an adaptive penalty mechanism. We link it to historical credit and behavior types. By decreasing the unit costs for manipulators, our model limits their impact on the final result.
3. Past models rarely use fairness as a direct mathematical constraint (Guo et al. [57]; Zhang et al. [55]). They usually handle it through external feedback rules. Zhang et al. [58] added fairness to individual utility, but this does not protect cooperative experts at the system level. Our model includes the Gini coefficient in the

optimization constraints. This sets a clear limit for cost allocation. It prevents cooperative individuals from taking on too much cost and ensures a balanced distribution.

To verify the effectiveness of the proposed method in the CRP, we conduct simulation experiments. These experiments compare various optimization models under the same initial conditions to evaluate their differences in total adjustment cost, consensus degree, and fairness degree. As shown in Figure 6, the performance of the seven different models is presented across these three metrics.




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

**Figure 6.** The total adjustment cost, consensus degree, and fairness degree in different optimization models [36,54–58]. (**a**) Total adjustment cost; (**b**) Consensus degree; (**c**) Fairness degree.

Figure 6a shows the total adjustment cost for each model. The single-objective models have the lowest costs. For example, Zhang et al. [55], Guo et al. [57], and Liu et al. [56] keep their costs between 14.2 and 14.9. In contrast, models with multiple objectives require more cost. The models from Shen et al. [36] and Zhao et al. [54] show costs above 17.0. The proposed model has a total cost of 15.31. This is slightly higher than the single-objective minimum cost models, but it is lower than the other multi-objective methods. Figure 6b displays the average consensus degree. Most of the single-objective minimum cost models reach a consensus level between 0.85 and 0.86. Zhao et al. [54] targets consensus directly, so it reaches a higher score of 0.891. The proposed model achieves the highest value at 0.894. This is higher than the single-objective minimum cost models. Figure 6c compares the fairness degree. Most of the single-objective minimum cost models score low on this metric. Their scores range from 0.71 to 0.78. This low score means the adjustment burden is not shared equally among the experts. Shen et al. [36] adds fairness as a target and improves the score to 0.848, while the proposed model scores 0.856. By using the Gini coefficient as a limit, it provides the most balanced cost allocation.

Overall, the data shows clearly that the single-objective minimum cost models minimize cost but result in lower fairness and lower consensus. The bi-objective models improve one specific area but perform worse in others. The proposed model balances all three targets. It accepts a small increase in total cost to achieve high consensus and ensure fairness for the experts.

### *8.3. Comparative Analysis of Different Behavior Identification and Management Mechanisms*

In this subsection, the dynamic identification and adaptive management mechanism for non-cooperative and manipulative behaviors designed in this study is compared with other related research. Comparisons are made in terms of identification strategies and management strategies for different types of consensus-breaking behaviors, as shown in Table 5.

Existing research on behavior identification primarily relies on static thresholds or cross-sectional data from a single consensus round. However, in dynamic environments, consensus-breaking behaviors can stem from either temporary strict preference retention

or deliberate long-term manipulation. These distinct motivations cannot be accurately distinguished using single-round data. Therefore, this study proposes the historical credit index to evaluate experts' long-term performance, providing a basis to differentiate short-term non-cooperation from persistent or manipulative intentions. Furthermore, in terms of specific identification metrics, existing methods often evaluate experts based on single dimensions, such as individual psychological traits or basic data-driven indicators. This fails to fully reflect the true adjustment intentions of experts. Compared to existing research, the dynamic identification framework proposed in this study uses the PFD to quantify adjustment amplitude and direction. By combining PFD with the historical credit index, support degree, and trust influence, this mechanism accurately identifies three specific types of behavior: short-term non-cooperative behavior, persistent non-cooperative behavior, and manipulative behavior.

**Table 5.** Comparison with existing methods for addressing non-cooperative and manipulative behaviors.

| Methods | Identification Strategy | Management Strategy |
|---|---|---|
| Liu et al. [59] | Based on experts' personality traits (agreeableness) and the calculated maximum acceptable adjustment amount, identifies 1 type of non-cooperative behavior. | Uses a maximum satisfaction-based feedback mechanism to maximize expert satisfaction and mitigate psychological resistance, avoiding direct punitive measures. |
| Fu et al. [37] | Based on the constructed adjustment willingness function and historical data-based uncertainty set analysis, identifies 1 type of non-cooperative behavior. | Employs a data-driven robust optimization framework to absorb consensus cost uncertainties caused by low adjustment willingness. |
| Zhong et al. [41] | Based on a multidimensional quantitative identification method using preference deviation and information asymmetry, identifies 2 types of behaviors: strategic manipulative behavior and manipulated behavior. | Removes the preference interference of manipulators from the network and deploys a personalized adjustment strategy to correct manipulated opinions. |
| Sun et al. [42] | Based on the monitoring of social trust status and dynamic analysis of power structures driven by self-esteem, identifies 1 type of strategic manipulative behavior. | Adjusts the group power structure dynamically via the maximum entropy principle to dilute the authority of manipulative high-trust nodes. |
| Peng et al. [60] | Based on the preference similarity of direct trust relationships and the analysis of social network topology, identifies 2 types of behaviors: non-cooperative behavior and leadership behavior. | Uses a behavior leader-based consensus filtering mechanism to aggregate opinions and exclude noise from non-cooperative edge nodes. |
| Proposed method | Based on the PFD and historical credit index, combined with support degree, and in-degree centrality, identifies 3 types of non-cooperative and manipulative behaviors, including short-term non-cooperative behavior, persistent non-cooperative behavior, and manipulative behavior. | Applies tiered penalties via graded trust decay and credit index reductions, with an influence-accelerated decay for manipulators. Dynamically links the historical credit index to unit adjustment costs, forcing low-credit experts to revise opinions. |

Most existing research manages consensus-breaking behaviors through independent feedback mechanisms, network structure adjustments, or cost optimization models. However, treating behavior penalties and consensus optimization as separate processes fails to establish a closed-loop system, and applying generalized responses ignores the different impacts of different consensus-breaking behaviors. Therefore, this study designs an adaptive management framework that implements targeted measures based on the label of

the identified behaviors. Specifically, it uses tiered penalties for the three non-cooperative and manipulative behaviors. These penalties include graded trust decay, trust redistribution, and reductions in the historical credit index. Meanwhile, cooperative experts receive credit rewards. More importantly, this study links the historical credit index directly to the unit adjustment cost coefficients in the tri-objective optimization model. This connection forces uncooperative experts to revise their opinions first to minimize the total system cost. As a result, the model successfully integrates behavior management with consensus optimization.

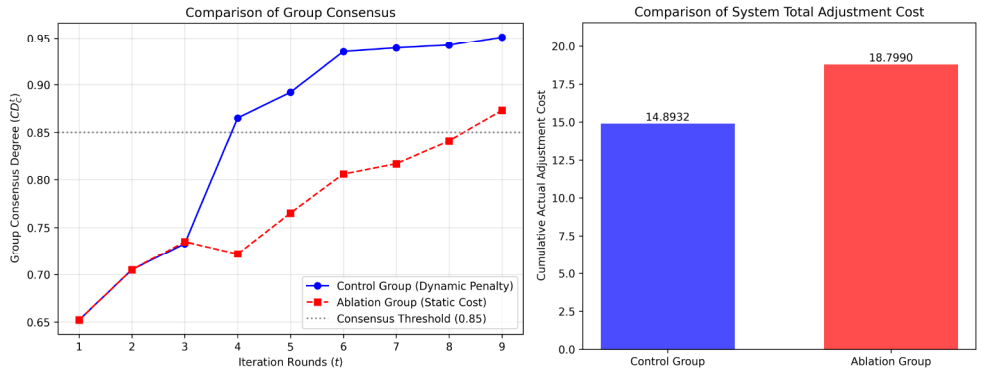
### *8.4. Ablation Analysis*

This section designs a comprehensive ablation study. It aims to quantify the impact of each component on the final performance metrics, thus providing solid empirical evidence of the effectiveness of the model architecture.

#### 8.4.1. Ablation Analysis of the Dynamic Unit Adjustment Cost Mechanism

In order to examine the contribution of the dynamic unit adjustment cost mechanism to the CRP, this section performs an ablation analysis to verify its necessity. As defined in Equation (40), it evaluates the impact of dynamically linking the historical credit index $C_u^t$ with the unit adjustment cost coefficient $\Psi_u^t$ on managing manipulative and non-cooperative behaviors.

To evaluate the effect of this mechanism on consensus efficiency and total adjustment cost, this study designs an experiment with two comparison groups based on the same initial opinion matrices and trust network structure. The control group uses the complete DBI-TOCM model, where the cost coefficient $\Psi_u^t$ of an expert decays when they are identified as non-cooperator or manipulator. The ablation group removes this dynamic penalty mechanism, locking the cost coefficient at its initial value ($\Psi_u^t = 1.0$) for all experts regardless of their behavior. We compare the performance of these two groups across varying iteration rounds, as illustrated in Figure 7.




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

**Figure 7.** Ablation analysis of the dynamic unit adjustment cost mechanism. (**a**) Comparison of group consensus; (**b**) Comparison of system total adjustment cost.

The results show that the dynamic penalty mechanism improves time efficiency and resource allocation. As shown in Figure 7a, the collective consensus degrees of both groups are identical during the first two rounds because experts have not accumulated enough behavioral data to trigger a penalty. A clear divergence appears in the third round. In the control group, the dynamic cost coefficients $\Psi_u^t$ for the identified non-cooperators and manipulators have dropped. This allows the optimization algorithm to overcome their resistance, causing the collective consensus degree $CD_C^t$ to rise up. It crosses the threshold of 0.85, reaches 0.865 in round 4 and converges above 0.95 in later rounds. Conversely,

https://doi.org/10.3390/electronics15143153

the ablation group treats the modification costs of all experts as equal. The algorithm struggles to adjust the non-cooperators' and manipulators' opinions so that it forces all experts to share the adjustment burden. Consequently, the CRP stalls, and the group only reaches 0.853 in round 9.

Moreover, Figure 7b demonstrates the impact of the mechanism on the actual adjustment cost. The control group records a cumulative adjustment cost of 14.8932, while the ablation group reaches 18.7990. Without the dynamic penalty, the static model cannot penalize manipulators effectively. To increase the consensus degree, the algorithm forces cooperative experts to make large and unnecessary opinion adjustments. In contrast, the control group uses the lowered unit cost coefficients to target and heavily modify the non-cooperators' and manipulators' opinions. This approach achieves global agreement at a much smaller physical cost.

These findings indicate that the dynamic unit adjustment cost mechanism is an essential component of the model. It targets behaviors that break consensus accurately. This prevents the system from generating unnecessary costs, ensuring an efficient CRP while protecting the interests of cooperative experts.

#### 8.4.2. Ablation Analysis of the Multi-Path Trust Aggregation Operator with Decayed Trust Propagation

An ablation study is conducted to validate the proposed trust aggregation algorithm. We compare the performance of the proposed method with variant models lacking key components. This comparison identifies the contribution of trust decay, path reliability, and path diversity to the quality of trust propagation and aggregation. The experiment is repeated on the same trust network to ensure consistency.

The specific method is as follows:

- $A_1$: The proposed method. It includes trust decay, path reliability, and path diversity. The path aggregation weight is set to $\Lambda_h = \frac{1}{2}\overline{RS}_h + \frac{1}{2}\overline{DS}_h$.
- $A_2$: The path diversity module is removed. The model relies only on path reliability for aggregation, where $\Lambda_h = \overline{RS}_h$. This group tests how topological redundancy affects trust evaluation.
- $A_3$: The path reliability module is removed. Aggregation depends solely on path diversity, setting $\Lambda_h = \overline{DS}_h$. This group examines the role of the "weakest link" principle in the propagation chain.
- $A_4$: The trust decay mechanism $w_h(i)$ is excluded. Propagation paths are assigned equal weights regardless of their topological distances. The structural aggregation modules are kept.
- $A_5$: A traditional baseline model. It removes both trust decay and structural aggregation. Multiple propagation paths are simply combined using an arithmetic mean.

This section introduces four core metrics to evaluate the quality of trust propagation. The definitions and evaluation criteria of these metrics are as follows:

- Propagation fidelity ($PF$): This metric measures the negative correlation between inferred indirect trust values and their topological distances. A higher value indicates that trust strength decays reasonably as the propagation path lengthens, preventing the overestimation of distant relationships.
- Topological independence ($TI$): This metric utilizes Shannon entropy to quantify the distribution uniformity of intermediate nodes across all valid propagation paths. A higher value implies that the trust inference relies on diverse structural routes, avoiding over-reliance on a few highly-connected hub nodes.
- Trust robustness ($TR$): This metric evaluates the stability of the global trust distribution after injecting a proportion of extreme low-trust links into the network. A higher

https://doi.org/10.3390/electronics15143153

score reflects smaller fluctuations in the overall assessment, demonstrating a stronger capability to isolate unreliable connections.

- Asymmetry preservation ratio ($APR$): This metric calculates the degree to which the directional differences of direct trust are maintained during inference. A higher ratio means the mechanism successfully preserves the inherent asymmetric nature of social networks, avoiding the inappropriate homogenization of directed relationships.

As shown in Figure 8, $A_1$ achieves the best overall performance among the five variants, indicating that trust decay, path reliability, and path diversity jointly improve the quality of trust propagation. Compared with $A_4$ and $A_5$, $A_1$ obtains higher propagation fidelity (PF), mainly because the decay mechanism limits the accumulated uncertainty of long paths and prevents distant indirect relations from retaining excessive trust values. Compared with $A_3$, $A_1$ shows stronger trust robustness (TR), suggesting that the weakest-link reliability score can suppress fragile paths containing low-trust edges. In addition, the higher topological independence (TI) of $A_1$ compared with $A_2$ and $A_5$ shows that the Jaccard diversity score reduces redundant information from overlapping paths and avoids repeated reliance on common intermediate experts. Overall, the decay factor controls long-path uncertainty, the weakest-link score suppresses fragile trust chains, and the Jaccard diversity score reduces structural redundancy. These mechanisms explain why the proposed operator produces more stable and topology-consistent trust values than the ablated variants.

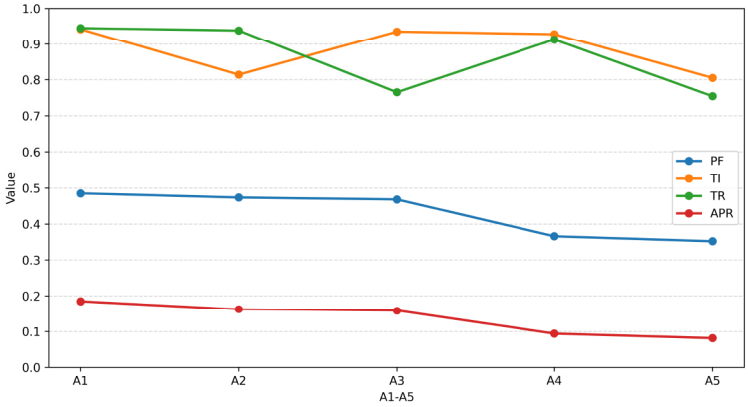



**Figure 8.** Ablation analysis of the multi-path trust aggregation operator with decayed trust propagation.

### 8.4.3. Ablation Analysis of the TOCM

To further verify the effectiveness of each key mechanism in the proposed TOCM, this section conducts a comprehensive ablation study. The first part evaluates the structural necessity of the adaptive boundary relaxation mechanism by comparing the complete model against a strictly constrained version. The second part validates the collaborative advantage of the multi-objective framework. This analysis compares the proposed model with three bi-objective versions where one objective function is removed.

(1) To verify the effectiveness of the adaptive boundary relaxation mechanism, TOCM-Strict is selected as the baseline model for comparative experiments. TOCM-Strict directly treats the initial global lower bound $\theta_{\min}$ and the consensus threshold $\epsilon$ as hard constraints within the NSGA-II optimization process. The aim is to reveal differences in solution feasibility under varying degrees of expert preservation tendency and consensus pressure.

As shown in Figure 9, TOCM-Strict assumes the optimization algorithm can naturally find feasible solutions within static predefined boundaries. It forces opinion modification under strict constraints. However, this structure is sensitive to initial

opinion divergence. The feasible region shrinks when the consensus threshold $\epsilon$ increases or when experts exhibit a higher opinion preservation tendency (larger $\theta_{\min}$). For example, at $\theta_{\min} = 0.6$, TOCM-Strict fails to output any feasible solution when $\epsilon \geq 0.90$. When $\theta_{\min}$ increases to 0.8, TOCM-Strict fails across all $\epsilon$ settings. This indicates that traditional single-stage hard constraint models are prone to consensus deadlock when handling real-world negotiation problems.

In contrast, the proposed TOCM introduces an asymmetric boundary relaxation mechanism. The algorithm identifies bottleneck experts based on dynamically changing consensus targets. It then attenuates the lower bounds only for these specific individuals. This ensures the existence of a feasible region before entering the multi-objective optimization phase. Experimental results show that the model maintains a solution success rate close to 1.0 under most conflict scenarios. The success rate decreases only under the strict condition of $\theta_{\min} = 0.8$ and $\epsilon = 0.95$. This aligns with the objective bounds of real-world negotiations. The results demonstrate that the two-stage mechanism resolves boundary constraint conflicts. It ensures the operational stability of the GDM model.

(2) The optimization framework in the proposed TOCM simultaneously considers three objectives: adjustment cost, consensus degree, and fairness degree. It aims to achieve more comprehensive decision support. By comparing the optimization results of the proposed model with three bi-objective ablation models, the collaborative advantage of the proposed framework is validated. Specifically, comparative analyses are conducted between the proposed model and each of the bi-objective models formed by removing one of the three objective functions: fairness, consensus, or cost.

Specifically, Figure 10a shows a positive correlation between consensus degree and total cost. Increasing the consensus level leads to higher adjustment costs. The bi-objective model ignoring fairness achieves high consensus with relative efficiency. However, it operates by assigning adjustments asymmetrically. This lowers group fairness and risks expert non-compliance. Figure 10b shows that higher fairness comes with a higher cost. The model ignoring consensus achieves higher fairness and maintains a lower total cost. However, it only satisfies the basic consensus threshold. It lacks sufficient group coordination to push opinions closer together, resulting in a fragile agreement. Figure 10c shows a negative correlation between consensus degree and fairness. Improving consensus requires sacrificing equitable cost distribution. The model ignoring cost achieves a mathematical balance between consensus and fairness. However, it risks uncontrolled cost growth and ignores the practical economic constraints of the experts. In summary, our method achieves a more comprehensive decision balance through multi-objective co-optimization, maintains solution diversity, and demonstrates practical potential.

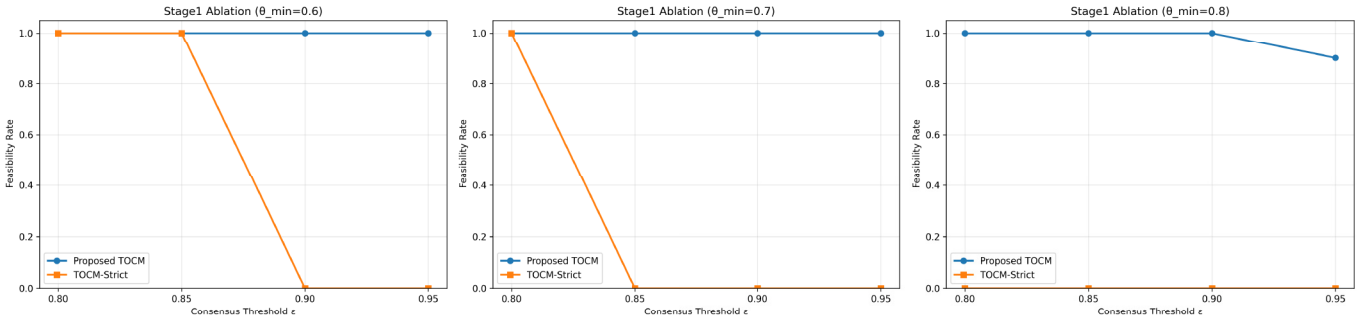
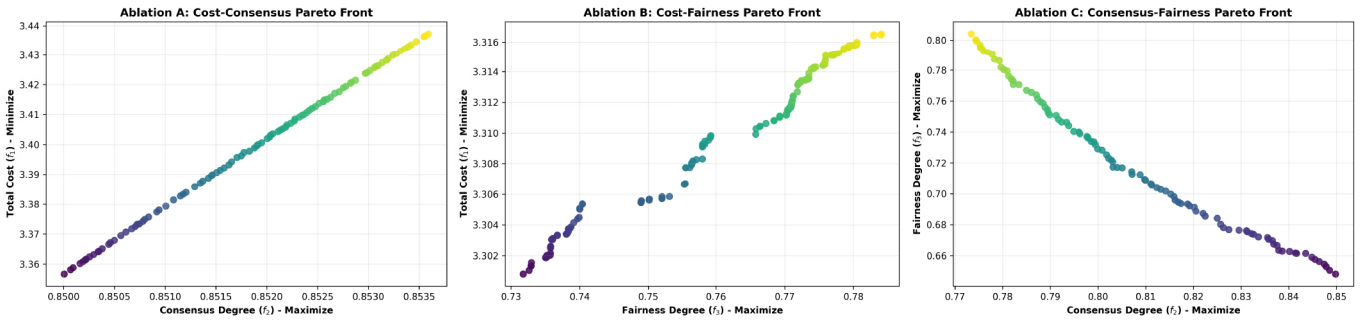



**Figure 9.** Comparison of solution feasibility rates between the proposed TOCM and TOCM-Strict under varying consensus thresholds. (**a**) $\theta_{\min} = 0.6$; (**b**) $\theta_{\min} = 0.7$; (**c**) $\theta_{\min} = 0.8$.




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

**Figure 10.** Pareto fronts of the three bi-objective ablation models.(**a**) total adjustment cost versus consensus degree; (**b**) total adjustment cost versus fairness degree; (**c**) consensus degree versus fairness degree.

### *8.5. The Hypothesis Testing Experiment*

To rigorously evaluate whether the performance differences between the proposed methods and the baseline models are statistically significant, we simulate $n$ independent SNGDM scenarios with varying initial opinion matrices and network topologies. Following the statistical evaluation approach used by Zhang et al. [28], we employ the paired samples $t$-test to verify these performance variations across the simulated scenarios. The paired samples $t$-test is an appropriate statistical inference tool for comparing the means of two distinct algorithms when applied to the identical set of experimental conditions. The formal procedure of the hypothesis testing is structured as follows:

First, the statistical hypotheses are defined:

- Null hypothesis ($H_0$): The mean difference in the evaluated performance metrics between the proposed algorithm and the baseline algorithm is zero.
- Alternative hypothesis ($H_1$): The mean difference in the evaluated performance metrics between the proposed algorithm and the baseline algorithm is significantly different from zero.

Let $n$ represent the total number of independently generated simulation scenarios. For a specific evaluation metric in the $i$-th scenario, let $X_{Pi}$ and $X_{Ci}$ denote the observed values obtained by the proposed method and a comparative baseline method, respectively. The paired difference is computed as $d_i = X_{Pi} - X_{Ci}$, where $i = 1, 2, \ldots, n$. Subsequently, the test statistic for the paired $t$-test is calculated using the following equation:

$$t = \frac{\bar{d}}{s_{\bar{d}}/\sqrt{n}}, \tag{43}$$

where $\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$ represents the sample mean of the paired differences; $s_d = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1}}$ denotes the sample standard deviation of these differences; and $n$ is the sample size.

To further quantify the magnitude of the performance difference, Cohen's $d_z$ for paired samples is reported as the standardized effect size:

$$d_z = \frac{\bar{d}}{s_d} = \frac{t}{\sqrt{n}}. \tag{44}$$

In addition, because multiple pairwise comparisons are conducted across different baseline methods and evaluation metrics, the raw $p$-values are adjusted using the Holm–Bonferroni procedure within each group of related comparisons. The adjusted $p$-value is reported together with the original $p$-value.

https://doi.org/10.3390/electronics15143153

In this experiment, all compared methods were evaluated under the same set of 50 randomly generated SNGDM scenarios. Therefore, the result of the proposed method and that of each baseline method in the same scenario form a paired observation, which justifies the use of the paired-samples $t$-test. Based on $n = 50$ repeated scenarios, the test statistic follows Student's $t$-distribution with $df = n - 1 = 49$. If the resulting $p$-value is less than the predetermined significance level ($\alpha = 0.05$), the null hypothesis $H_0$ is rejected.

Since each paired comparison was based on 50 repeated scenarios, the paired-samples $t$-test was used as the main inferential test. The test was applied to the paired differences between the proposed method and each baseline method under identical experimental scenarios. Considering the moderate sample size, the $t$-test is relatively robust to mild deviations from normality. Therefore, the statistical conclusions were interpreted together with effect sizes and multiple-comparison-adjusted $p$-values, rather than relying only on raw $p$-values.

#### 8.5.1. The Hypothesis Testing for the Trust Propagation Algorithm

To verify the effectiveness of the proposed multi-path decayed trust aggregation method, this section conducts a hypothesis testing experiment comparing it with existing approaches. This section uses propagation fidelity (PF), topological independence (TI), and trust robustness (TR) as evaluation metrics.

As shown in Table 6, for all three metrics, the raw $p$-values of the paired $t$-tests between the proposed method and all baseline methods are less than 0.05. After applying the Holm–Bonferroni correction, all adjusted $p$-values remain below 0.05. In addition, the absolute values of Cohen's $d_z$ range from 0.341 to 0.648, indicating a clear standardized difference between the proposed method and the baseline methods. According to the hypothesis-testing criteria, the null hypothesis is rejected and the alternative hypothesis is accepted. The statistical conclusion confirms that the proposed method has a significant advantage over traditional approaches in maintaining trust propagation fidelity and structural robustness.

**Table 6.** The paired $t$-test results for the trust propagation algorithm.

| Comparison Methods | Evaluation Metrics | $t$-Values | $p$-Value | Adjusted $p$-Value | Cohen's $d_z$ | Significance Levels ($p_{adj} < 0.05$) |
|---|---|---|---|---|---|---|
| Yuan et al. [25] | PF | 4.251 | <0.001 | <0.001 | 0.601 | yes |
| | TI | 4.582 | <0.001 | <0.001 | 0.648 | yes |
| | TR | 3.914 | <0.001 | 0.00196 | 0.554 | yes |
| Zhao et al. [54] | PF | 3.342 | 0.00160 | 0.00799 | 0.473 | yes |
| | TI | 3.651 | <0.001 | 0.00380 | 0.516 | yes |
| | TR | 3.125 | 0.00299 | 0.01194 | 0.442 | yes |
| Gai et al. [27] | PF | 2.543 | 0.01420 | 0.02840 | 0.360 | yes |
| | TI | 2.781 | 0.00767 | 0.02301 | 0.393 | yes |
| | TR | 2.412 | 0.01966 | 0.02840 | 0.341 | yes |

#### 8.5.2. The Hypothesis Testing for the TOCM

To verify the performance advantage of the proposed TOCM, this section selects two core metrics, total adjustment cost and fairness degree, and compares it with existing methods.

As shown in Table 7, the raw $p$-values for all comparison experiments are less than 0.05. After Holm–Bonferroni correction, most comparisons remain statistically significant. Specifically, compared with single-objective models, the proposed model shows significant improvement in fairness, while the cost differences are not significant after correction. This result is consistent with the design of the proposed TOCM, which does not pursue cost

minimization alone but jointly considers adjustment cost, consensus degree, and fairness. Conversely, compared with multi-objective models, the proposed model demonstrates a significant reduction in total adjustment cost while further enhancing fairness. Cohen's $d_z$ values also show non-negligible standardized differences across the comparisons. This indicates that the proposed consensus optimization model achieves a statistical advantage in balancing cost efficiency and cost fairness.

**Table 7.** The paired *t*-test results for the TOCM.

| Comparison Methods | Evaluation Metrics | *t*-Values | *p*-Value | Adjusted *p*-Value | Cohen's $d_z$ | Significance Levels ($p_{adj} < 0.05$) |
|---|---|---|---|---|---|---|
| Zhang et al. [55] | Cost | 2.431 | 0.01876 | 0.05629 | 0.344 | no |
| | Fairness | 5.892 | <0.001 | <0.001 | 0.833 | yes |
| Liu et al. [56] | Cost | 2.215 | 0.03144 | 0.06288 | 0.313 | no |
| | Fairness | 5.341 | <0.001 | <0.001 | 0.755 | yes |
| Guo et al. [57] | Cost | 2.154 | 0.03619 | 0.06288 | 0.305 | no |
| | Fairness | 4.925 | <0.001 | <0.001 | 0.697 | yes |
| Shen et al. [36] | Cost | −3.672 | <0.001 | 0.00476 | −0.519 | yes |
| | Fairness | 2.651 | 0.01078 | 0.04311 | 0.375 | yes |
| Zhang et al. [58] | Cost | −3.124 | 0.00299 | 0.01797 | −0.442 | yes |
| | Fairness | 2.892 | 0.00569 | 0.02847 | 0.409 | yes |
| Zhao et al. [54] | Cost | −4.152 | <0.001 | 0.00118 | −0.587 | yes |
| | Fairness | 3.254 | 0.00207 | 0.01446 | 0.460 | yes |

#### 8.5.3. The Hypothesis Testing for the Dynamic Behavior Identification and Management Mechanism

To verify the overall performance of the dynamic behavior identification and management mechanism, this section compares it with existing strategies, focusing on two efficiency metrics: consensus reaching rounds and total adjustment cost.

As shown in Table 8, the proposed mechanism demonstrates statistically significant performance advantages over all comparisons ($p < 0.05$). According to the hypothesis-testing criteria, we reject the null hypothesis $H_0$ and accept the alternative hypothesis $H_1$. This statistical conclusion confirms that the advantages of the proposed mechanism in accelerating consensus and controlling costs are stable outcomes.

**Table 8.** The paired *t*-test results for the dynamic behavior identification and management mechanism.

| Comparison Methods | Evaluation Metrics | *t*-Values | *p*-Value | Adjusted *p*-Value | Cohen's $d_z$ | Significance Levels ($p_{adj} < 0.05$) |
|---|---|---|---|---|---|---|
| Liu et al. [59] | Cost | −4.321 | <0.001 | <0.001 | −0.611 | yes |
| | Rounds | −3.981 | <0.001 | 0.00204 | −0.563 | yes |
| Fu et al. [37] | Cost | −3.815 | <0.001 | 0.00306 | −0.540 | yes |
| | Rounds | −3.542 | <0.001 | 0.00530 | −0.501 | yes |
| Zhong et al. [41] | Cost | −3.104 | 0.00317 | 0.01267 | −0.439 | yes |
| | Rounds | −2.761 | 0.00809 | 0.01617 | −0.390 | yes |
| Sun et al. [42] | Cost | −3.652 | <0.001 | 0.00443 | −0.516 | yes |
| | Rounds | −3.214 | 0.00232 | 0.01159 | −0.455 | yes |
| Peng et al. [60] | Cost | −2.932 | 0.00511 | 0.01532 | −0.415 | yes |
| | Rounds | −2.451 | 0.01786 | 0.01786 | −0.347 | yes |

## 9. Conclusions

This paper proposes a dynamic consensus optimization approach for SNGDM. First, a multi-path decayed trust aggregation operator is designedto measure path reliability and structural diversity, thereby preventing trust distortion. Second, a tri-objective optimization model is constructed to find optimal adjustment strategies by balancing total adjustment cost, consensus degree, and cost fairness. Traditional consensus mechanisms often rely on static penalty rules. They usually separate behavior management from the actual optimization process. Compared with these traditional methods, our approach creates an adaptive, closed-loop system. It links the historical credit index of each expert to their unit cost coefficient. When experts display non-cooperative or manipulative behaviors, the system automatically reduces their trust weights and unit cost coefficients. This mechanism forces violators to adjust their opinions, which effectively prevents them from exploiting the network structure.

The results of the analysis reveal two key findings. First, the sensitivity analysis demonstrates a U-shaped relationship between the tolerance threshold and the total adjustment cost. Identifying the optimal threshold range balances the accuracy of behavior identification with the overall adjustment cost. A threshold set too low allows non-cooperators to bypass detection by making minor concessions. In contrast, a threshold set too high mechanically penalizes cooperative experts. This misjudgment damages the underlying trust network and forces the system to require higher costs to reach an agreement. Second, the ablation study confirms the necessity of the dynamic unit cost coefficient mechanism. By dynamically lowering the cost coefficients of identified manipulators, the optimization algorithm prioritizes modifying their opinions. This prevents the CRP from stalling. It also protects cooperative experts from taking on unfair adjustment burdens. The proposed method is illustrated through a case study on the procurement of a smart city medical resource scheduling platform. In this setting, the decision panel includes experts from different professional backgrounds. Their priorities, information bases, and network positions are not identical. The proposed framework can be used to identify strategic manipulation, manage non-cooperative behavior, and support a more balanced CRP.

Beyond this healthcare procurement case, the proposed DBI-TOCM can be used in other SNGDM contexts. One application is public procurement and smart city governance, where government departments, technical experts, financial authorities, and service providers jointly evaluate complex alternatives under incomplete trust relationships. It can also be used in emergency management and public health resource allocation. In these situations, experts need to reach agreement under uncertainty, time pressure, and institutional conflicts. Another application is environmental governance, urban planning, and infrastructure project evaluation, where stakeholders often differ in their views on cost, risk, fairness, and long-term public benefits. The method can also be extended to enterprise decision-making, such as supplier selection, project portfolio evaluation, and technology adoption. In such cases, decision makers are often connected through organizational trust networks and face different adjustment costs. These application areas share several common features: uncertain expert preferences, incomplete trust relationships, and the possibility that non-cooperative or manipulative behaviors may affect the CRP.

Despite these practical advantages, the proposed method has specific limitations. First, the current framework assumes that experts provide complete opinion matrices. In actual decision-making, experts may lack the domain knowledge to evaluate every alternative. They often provide incomplete information. This missing data affects the initial trust aggregation and the subsequent optimization process. Second, the dynamic identification module focuses entirely on individual actions. It detects a single non-cooperator or manipulator effectively. However, it contains a blind spot for group collusion. A subgroup of experts could coordinate

to assign biased scores and artificially inflate their mutual trust. This type of coordinated behavior can bypass the current single-node detection rules. Future research should address these gaps. Specifically, methods can be developed to handle incomplete evaluation information and measure the reliability of experts. Furthermore, the behavior identification framework can be expanded from the individual level to the subgroup level. By incorporating network cluster analysis and monitoring coordinated opinion adjustments, future models can be equipped to detect and manage collective non-cooperation and manipulation.

**Author Contributions:** Conceptualization, W.D., W.C. and C.Z.; Methodology, W.D. and W.C.; Validation, W.D., W.C., H.Z., J.P. and C.Z.; Data curation, W.D., H.Z. and J.P.; Writing—original draft, W.D. and W.C.; Writing—review and editing, C.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Central Government Guides Local Science and Technology Innovation (YDZJSX2024D015), the Science and Technology Cooperation and Exchange Special Projects of Shanxi (202504041101007), the Top Young Talents of Shanxi "Three Jin" Talents Program (SJYC2025334) and the Wenying Young Scholars of Shanxi University.

**Data Availability Statement:** The data will be made available upon request.

**Conflicts of Interest:** Authors Hui Zhang and Jinchun Peng are employed by the company iGALAXY Information 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 potential conflicts of interest.

## Appendix A

*Appendix A.1*

Initial opinion matrices of each expert.

$$D_1^1 = \begin{bmatrix} (0.88,0.24) & (0.82,0.30) & (0.86,0.26) \\ (0.60,0.54) & (0.64,0.48) & (0.62,0.50) \\ (0.52,0.62) & (0.50,0.64) & (0.58,0.56) \\ (0.36,0.76) & (0.66,0.46) & (0.38,0.74) \end{bmatrix} \quad D_2^1 = \begin{bmatrix} (0.94,0.16) & (0.78,0.34) & (0.90,0.20) \\ (0.30,0.82) & (0.52,0.60) & (0.38,0.76) \\ (0.66,0.48) & (0.48,0.66) & (0.72,0.40) \\ (0.22,0.88) & (0.76,0.34) & (0.28,0.84) \end{bmatrix}$$

$$D_3^1 = \begin{bmatrix} (0.56,0.60) & (0.34,0.82) & (0.48,0.66) \\ (0.44,0.70) & (0.60,0.52) & (0.42,0.72) \\ (0.52,0.62) & (0.58,0.56) & (0.56,0.60) \\ (0.82,0.30) & (0.90,0.18) & (0.84,0.28) \end{bmatrix} \quad D_4^1 = \begin{bmatrix} (0.78,0.34) & (0.30,0.82) & (0.64,0.50) \\ (0.24,0.86) & (0.52,0.60) & (0.36,0.78) \\ (0.92,0.18) & (0.86,0.24) & (0.94,0.16) \\ (0.20,0.90) & (0.72,0.38) & (0.28,0.84) \end{bmatrix}$$

$$D_5^1 = \begin{bmatrix} (0.60,0.54) & (0.28,0.84) & (0.62,0.50) \\ (0.24,0.86) & (0.50,0.62) & (0.34,0.78) \\ (0.88,0.22) & (0.84,0.26) & (0.95,0.15) \\ (0.18,0.92) & (0.70,0.40) & (0.26,0.86) \end{bmatrix} \quad D_6^1 = \begin{bmatrix} (0.86,0.24) & (0.42,0.72) & (0.82,0.30) \\ (0.38,0.76) & (0.64,0.48) & (0.46,0.70) \\ (0.90,0.20) & (0.78,0.34) & (0.92,0.18) \\ (0.24,0.86) & (0.70,0.40) & (0.30,0.82) \end{bmatrix}$$

*Appendix A.2*

Opinion matrices of each expert in round 2.

$$D_1^2 = \begin{bmatrix} (0.93,0.19) & (0.86,0.26) & (0.91,0.21) \\ (0.57,0.57) & (0.62,0.50) & (0.60,0.53) \\ (0.52,0.63) & (0.49,0.65) & (0.58,0.57) \\ (0.35,0.78) & (0.65,0.47) & (0.37,0.76) \end{bmatrix} \quad D_2^2 = \begin{bmatrix} (0.92,0.18) & (0.75,0.38) & (0.88,0.23) \\ (0.34,0.80) & (0.54,0.59) & (0.40,0.75) \\ (0.71,0.44) & (0.57,0.59) & (0.76,0.37) \\ (0.33,0.85) & (0.76,0.35) & (0.37,0.81) \end{bmatrix}$$

$$D_3^2 = \begin{bmatrix} (0.55,0.61) & (0.33,0.83) & (0.47,0.67) \\ (0.42,0.72) & (0.58,0.54) & (0.40,0.74) \\ (0.51,0.63) & (0.57,0.57) & (0.55,0.61) \\ (0.86,0.26) & (0.94,0.14) & (0.88,0.24) \end{bmatrix} \quad D_4^2 = \begin{bmatrix} (0.80,0.33) & (0.44,0.75) & (0.69,0.46) \\ (0.31,0.83) & (0.54,0.59) & (0.39,0.76) \\ (0.91,0.20) & (0.85,0.27) & (0.93,0.18) \\ (0.32,0.86) & (0.73,0.38) & (0.37,0.81) \end{bmatrix}$$

$$D_5^2 = \begin{bmatrix} (0.66,0.50) & (0.36,0.80) & (0.65,0.48) \\ (0.27,0.85) & (0.51,0.61) & (0.36,0.77) \\ (0.87,0.23) & (0.83,0.28) & (0.94,0.16) \\ (0.29,0.89) & (0.71,0.39) & (0.33,0.83) \end{bmatrix} \quad D_6^2 = \begin{bmatrix} (0.85,0.26) & (0.47,0.70) & (0.81,0.32) \\ (0.39,0.76) & (0.62,0.50) & (0.46,0.70) \\ (0.91,0.20) & (0.79,0.34) & (0.93,0.18) \\ (0.33,0.84) & (0.71,0.40) & (0.37,0.80) \end{bmatrix}$$

Opinion matrices of each expert in round 3.

$$D_1^3 = \begin{bmatrix} (0.94,0.18) & (0.87,0.27) & (0.92,0.21) \\ (0.55,0.60) & (0.61,0.52) & (0.57,0.56) \\ (0.58,0.59) & (0.53,0.63) & (0.63,0.53) \\ (0.38,0.77) & (0.66,0.46) & (0.40,0.75) \end{bmatrix} \quad D_2^3 = \begin{bmatrix} (0.91,0.20) & (0.74,0.41) & (0.87,0.25) \\ (0.36,0.79) & (0.54,0.58) & (0.42,0.73) \\ (0.75,0.41) & (0.62,0.55) & (0.79,0.35) \\ (0.40,0.81) & (0.76,0.35) & (0.43,0.78) \end{bmatrix}$$

$$D_3^3 = \begin{bmatrix} (0.54,0.62) & (0.33,0.84) & (0.47,0.68) \\ (0.40,0.74) & (0.56,0.56) & (0.39,0.75) \\ (0.50,0.64) & (0.55,0.59) & (0.55,0.62) \\ (0.90,0.22) & (0.99,0.10) & (0.92,0.20) \end{bmatrix} \quad D_4^3 = \begin{bmatrix} (0.81,0.32) & (0.50,0.71) & (0.71,0.44) \\ (0.33,0.82) & (0.54,0.58) & (0.40,0.75) \\ (0.91,0.21) & (0.84,0.28) & (0.93,0.19) \\ (0.38,0.84) & (0.73,0.37) & (0.42,0.79) \end{bmatrix}$$

$$D_5^3 = \begin{bmatrix} (0.66,0.49) & (0.38,0.80) & (0.66,0.47) \\ (0.28,0.84) & (0.52,0.61) & (0.36,0.77) \\ (0.87,0.23) & (0.83,0.28) & (0.94,0.16) \\ (0.30,0.89) & (0.71,0.39) & (0.35,0.83) \end{bmatrix} \quad D_6^3 = \begin{bmatrix} (0.84,0.27) & (0.52,0.67) & (0.80,0.33) \\ (0.39,0.76) & (0.61,0.51) & (0.45,0.70) \\ (0.91,0.21) & (0.80,0.33) & (0.94,0.18) \\ (0.39,0.81) & (0.72,0.39) & (0.43,0.77) \end{bmatrix}$$

Opinion matrices of each expert in round 4.

$$D_1^4 = \begin{bmatrix} (0.94,0.19) & (0.85,0.30) & (0.91,0.22) \\ (0.53,0.63) & (0.60,0.52) & (0.55,0.58) \\ (0.66,0.53) & (0.59,0.58) & (0.71,0.47) \\ (0.45,0.76) & (0.71,0.43) & (0.47,0.73) \end{bmatrix} \quad D_2^4 = \begin{bmatrix} (0.90,0.22) & (0.72,0.44) & (0.85,0.27) \\ (0.37,0.78) & (0.55,0.58) & (0.42,0.73) \\ (0.77,0.39) & (0.65,0.51) & (0.82,0.33) \\ (0.46,0.79) & (0.77,0.34) & (0.49,0.75) \end{bmatrix}$$

$$D_3^4 = \begin{bmatrix} (0.54,0.63) & (0.33,0.85) & (0.47,0.69) \\ (0.39,0.75) & (0.55,0.57) & (0.37,0.77) \\ (0.52,0.63) & (0.56,0.58) & (0.56,0.61) \\ (0.95,0.18) & (1.00,0.05) & (0.97,0.16) \end{bmatrix} \quad D_4^4 = \begin{bmatrix} (0.81,0.32) & (0.53,0.69) & (0.73,0.43) \\ (0.34,0.81) & (0.54,0.58) & (0.41,0.74) \\ (0.91,0.22) & (0.84,0.29) & (0.93,0.19) \\ (0.42,0.82) & (0.75,0.37) & (0.45,0.77) \end{bmatrix}$$

$$D_5^4 = \begin{bmatrix} (0.67,0.49) & (0.38,0.80) & (0.66,0.47) \\ (0.28,0.84) & (0.52,0.61) & (0.36,0.77) \\ (0.87,0.23) & (0.83,0.28) & (0.94,0.16) \\ (0.30,0.89) & (0.71,0.39) & (0.35,0.83) \end{bmatrix} \quad D_6^4 = \begin{bmatrix} (0.88,0.68) & (0.54,0.82) & (0.86,0.71) \\ (0.93,0.04) & (0.92,0.04) & (0.99,0.11) \\ (0.89,0.67) & (0.96,0.50) & (0.33,0.98) \\ (0.59,0.73) & (0.40,0.01) & (0.03,0.65) \end{bmatrix}$$

https://doi.org/10.3390/electronics15143153

Opinion matrices of each expert in round 5.

$$D_1^5 = \begin{bmatrix} (0.93, 0.23) & (0.82, 0.36) & (0.90, 0.26) \\ (0.61, 0.51) & (0.66, 0.43) & (0.70, 0.53) \\ (0.73, 0.50) & (0.70, 0.55) & (0.73, 0.47) \\ (0.51, 0.74) & (0.77, 0.30) & (0.52, 0.70) \end{bmatrix} \quad D_2^5 = \begin{bmatrix} (0.89, 0.25) & (0.71, 0.47) & (0.85, 0.30) \\ (0.53, 0.63) & (0.62, 0.48) & (0.63, 0.64) \\ (0.80, 0.39) & (0.73, 0.49) & (0.82, 0.35) \\ (0.50, 0.77) & (0.80, 0.28) & (0.52, 0.73) \end{bmatrix}$$

$$D_3^5 = \begin{bmatrix} (0.54, 0.63) & (0.33, 0.86) & (0.47, 0.69) \\ (0.44, 0.69) & (0.58, 0.53) & (0.46, 0.72) \\ (0.52, 0.64) & (0.56, 0.60) & (0.56, 0.62) \\ (0.98, 0.15) & (1.00, 0.01) & (1.00, 0.13) \end{bmatrix} \quad D_4^5 = \begin{bmatrix} (0.83, 0.33) & (0.57, 0.67) & (0.76, 0.43) \\ (0.51, 0.66) & (0.61, 0.49) & (0.61, 0.65) \\ (0.90, 0.25) & (0.85, 0.31) & (0.92, 0.22) \\ (0.47, 0.79) & (0.77, 0.30) & (0.49, 0.74) \end{bmatrix}$$

$$D_5^5 = \begin{bmatrix} (0.73, 0.47) & (0.46, 0.76) & (0.71, 0.47) \\ (0.51, 0.65) & (0.61, 0.49) & (0.62, 0.65) \\ (0.87, 0.27) & (0.84, 0.31) & (0.92, 0.21) \\ (0.43, 0.84) & (0.76, 0.29) & (0.44, 0.78) \end{bmatrix} \quad D_6^5 = \begin{bmatrix} (0.42, 0.73) & (0.38, 0.75) & (0.42, 0.74) \\ (0.95, 0.16) & (0.95, 0.16) & (0.95, 0.17) \\ (0.42, 0.73) & (0.43, 0.71) & (0.36, 0.77) \\ (0.39, 0.74) & (0.36, 0.66) & (0.32, 0.73) \end{bmatrix}$$

Opinion matrices of each expert in round 6.

$$D_1^6 = \begin{bmatrix} (0.92, 0.25) & (0.80, 0.39) & (0.89, 0.29) \\ (0.67, 0.49) & (0.70, 0.41) & (0.74, 0.50) \\ (0.74, 0.49) & (0.70, 0.54) & (0.74, 0.46) \\ (0.51, 0.74) & (0.77, 0.31) & (0.53, 0.71) \end{bmatrix} \quad D_2^6 = \begin{bmatrix} (0.88, 0.26) & (0.70, 0.48) & (0.84, 0.31) \\ (0.57, 0.61) & (0.64, 0.47) & (0.66, 0.62) \\ (0.80, 0.39) & (0.73, 0.49) & (0.82, 0.35) \\ (0.50, 0.77) & (0.79, 0.28) & (0.52, 0.73) \end{bmatrix}$$

$$D_3^6 = \begin{bmatrix} (0.53, 0.65) & (0.32, 0.87) & (0.46, 0.71) \\ (0.46, 0.69) & (0.59, 0.53) & (0.48, 0.72) \\ (0.51, 0.66) & (0.54, 0.61) & (0.55, 0.63) \\ (1.00, 0.11) & (1.00, 0.00) & (1.00, 0.09) \end{bmatrix} \quad D_4^6 = \begin{bmatrix} (0.82, 0.34) & (0.58, 0.65) & (0.76, 0.43) \\ (0.58, 0.61) & (0.66, 0.47) & (0.66, 0.61) \\ (0.90, 0.26) & (0.84, 0.33) & (0.92, 0.24) \\ (0.48, 0.79) & (0.77, 0.31) & (0.50, 0.74) \end{bmatrix}$$

$$D_5^6 = \begin{bmatrix} (0.74, 0.46) & (0.50, 0.73) & (0.71, 0.47) \\ (0.61, 0.59) & (0.67, 0.45) & (0.69, 0.60) \\ (0.86, 0.29) & (0.82, 0.34) & (0.91, 0.23) \\ (0.45, 0.82) & (0.77, 0.30) & (0.47, 0.77) \end{bmatrix} \quad D_6^6 = \begin{bmatrix} (0.37, 0.74) & (0.36, 0.74) & (0.37, 0.74) \\ (0.95, 0.18) & (0.95, 0.18) & (0.95, 0.18) \\ (0.37, 0.74) & (0.37, 0.74) & (0.36, 0.74) \\ (0.36, 0.74) & (0.36, 0.73) & (0.36, 0.74) \end{bmatrix}$$

Opinion matrices of each expert in round 7.

$$D_1^7 = \begin{bmatrix} (0.93, 0.24) & (0.81, 0.39) & (0.89, 0.28) \\ (0.67, 0.49) & (0.70, 0.41) & (0.73, 0.50) \\ (0.73, 0.49) & (0.70, 0.54) & (0.74, 0.46) \\ (0.51, 0.74) & (0.97, 0.05) & (0.53, 0.71) \end{bmatrix} \quad D_2^7 = \begin{bmatrix} (0.88, 0.26) & (0.70, 0.48) & (0.84, 0.32) \\ (0.57, 0.61) & (0.64, 0.47) & (0.65, 0.62) \\ (0.81, 0.38) & (0.74, 0.48) & (0.83, 0.34) \\ (0.50, 0.77) & (0.98, 0.02) & (0.52, 0.73) \end{bmatrix}$$

$$D_3^7 = \begin{bmatrix} (0.52, 0.66) & (0.31, 0.89) & (0.45, 0.72) \\ (0.48, 0.68) & (0.59, 0.52) & (0.49, 0.71) \\ (0.50, 0.67) & (0.53, 0.63) & (0.53, 0.65) \\ (1.00, 0.07) & (1.00, 0.00) & (1.00, 0.05) \end{bmatrix} \quad D_4^7 = \begin{bmatrix} (0.82, 0.34) & (0.58, 0.66) & (0.75, 0.43) \\ (0.58, 0.62) & (0.65, 0.47) & (0.66, 0.62) \\ (0.91, 0.25) & (0.85, 0.32) & (0.93, 0.23) \\ (0.48, 0.79) & (0.97, 0.04) & (0.50, 0.74) \end{bmatrix}$$

$$D_5^7 = \begin{bmatrix} (0.74, 0.46) & (0.50, 0.73) & (0.71, 0.47) \\ (0.61, 0.59) & (0.67, 0.45) & (0.69, 0.60) \\ (0.86, 0.29) & (0.82, 0.34) & (0.91, 0.23) \\ (0.45, 0.82) & (0.97, 0.03) & (0.47, 0.77) \end{bmatrix} \quad D_6^7 = \begin{bmatrix} (0.36, 0.74) & (0.36, 0.74) & (0.36, 0.74) \\ (0.95, 0.18) & (0.95, 0.18) & (0.95, 0.18) \\ (0.36, 0.74) & (0.36, 0.74) & (0.36, 0.74) \\ (0.36, 0.74) & (0.36, 0.74) & (0.36, 0.74) \end{bmatrix}$$

https://doi.org/10.3390/electronics15143153

*Appendix A.3*

The initial trust.

$$S = \begin{bmatrix} 1 & 0.62 & 0.46 & 0.58 & 0 & 0.72 \\ 0.56 & 1 & 0.42 & 0.54 & 0 & 0.66 \\ 0.48 & 0.36 & 1 & 0 & 0.30 & 0.58 \\ 0.64 & 0.52 & 0.44 & 1 & 0 & 0.76 \\ 0 & 0.40 & 0.34 & 0.50 & 1 & 0.60 \\ 0.58 & 0 & 0.38 & 0.62 & 0 & 1 \end{bmatrix}$$

*Appendix A.4*

The complete trust matrices for 7 rounds of CRP.

$$S^1 = \begin{bmatrix} 1.0000 & 0.6200 & 0.4600 & 0.5800 & 0.0642 & 0.7200 \\ 0.5600 & 1.0000 & 0.4200 & 0.5400 & 0.0574 & 0.6600 \\ 0.4800 & 0.3600 & 1.0000 & 0.1923 & 0.3000 & 0.5800 \\ 0.6400 & 0.5200 & 0.4400 & 1.0000 & 0.0608 & 0.7600 \\ 0.1863 & 0.4000 & 0.3400 & 0.5000 & 1.0000 & 0.6000 \\ 0.5800 & 0.1923 & 0.3800 & 0.6200 & 0.0508 & 1.0000 \end{bmatrix}$$

$$S^2 = \begin{bmatrix} 1.0000 & 0.6200 & 0.4600 & 0.5800 & 0.0642 & 0.7200 \\ 0.5600 & 1.0000 & 0.4200 & 0.5400 & 0.0574 & 0.6600 \\ 0.4800 & 0.3600 & 1.0000 & 0.1923 & 0.3000 & 0.5800 \\ 0.6400 & 0.5200 & 0.4400 & 1.0000 & 0.0608 & 0.7600 \\ 0.1863 & 0.4000 & 0.3400 & 0.5000 & 1.0000 & 0.6000 \\ 0.5800 & 0.1923 & 0.3800 & 0.6200 & 0.0508 & 1.0000 \end{bmatrix}$$

$$S^3 = \begin{bmatrix} 1.0000 & 0.6337 & 0.4162 & 0.5928 & 0.0657 & 0.7359 \\ 0.5723 & 1.0000 & 0.3800 & 0.5519 & 0.0586 & 0.6745 \\ 0.4800 & 0.3600 & 1.0000 & 0.1923 & 0.3000 & 0.5800 \\ 0.6535 & 0.5310 & 0.3981 & 1.0000 & 0.0621 & 0.7761 \\ 0.1899 & 0.4077 & 0.3076 & 0.5096 & 1.0000 & 0.6115 \\ 0.5945 & 0.1972 & 0.3138 & 0.6355 & 0.0521 & 1.0000 \end{bmatrix}$$

$$S^4 = \begin{bmatrix} 1.0000 & 0.6561 & 0.2284 & 0.6138 & 0.0680 & 0.8779 \\ 0.5923 & 1.0000 & 0.2086 & 0.5712 & 0.0607 & 0.8047 \\ 0.4397 & 0.3298 & 1.0000 & 0.1762 & 0.2748 & 0.6919 \\ 0.6692 & 0.5437 & 0.2185 & 1.0000 & 0.0635 & 0.9258 \\ 0.1934 & 0.4153 & 0.1688 & 0.5192 & 1.0000 & 0.7295 \\ 0.6569 & 0.2178 & 0.1887 & 0.7022 & 0.0575 & 1.0000 \end{bmatrix}$$

$$S^5 = \begin{bmatrix} 1.0000 & 0.7062 & 0.0592 & 0.6607 & 0.0732 & 0.9449 \\ 0.6374 & 1.0000 & 0.0541 & 0.6146 & 0.0653 & 0.8659 \\ 0.4397 & 0.3298 & 1.0000 & 0.1762 & 0.2748 & 0.6919 \\ 0.7184 & 0.5837 & 0.0566 & 1.0000 & 0.0682 & 0.9939 \\ 0.2065 & 0.4433 & 0.0438 & 0.5541 & 1.0000 & 0.7786 \\ 0.7131 & 0.2365 & 0.0489 & 0.7622 & 0.0625 & 1.0000 \end{bmatrix}$$

$$S^6 = \begin{bmatrix} 1.0000 & 1.0000 & 0.0132 & 0.9492 & 0.1051 & 0.3619 \\ 0.9162 & 1.0000 & 0.0121 & 0.8835 & 0.0939 & 0.3317 \\ 0.5935 & 0.4451 & 1.0000 & 0.2378 & 0.3709 & 0.2650 \\ 1.0000 & 0.8636 & 0.0126 & 1.0000 & 0.1009 & 0.3807 \\ 0.2947 & 0.6327 & 0.0098 & 0.7909 & 1.0000 & 0.2982 \\ 0.7283 & 0.2415 & 0.0109 & 0.7786 & 0.0638 & 1.0000 \end{bmatrix}$$

$$S^7 = \begin{bmatrix} 1.0000 & 1.0000 & 0.0029 & 1.0000 & 0.1201 & 0.0808 \\ 1.0000 & 1.0000 & 0.0027 & 1.0000 & 0.1071 & 0.0740 \\ 0.6677 & 0.5007 & 1.0000 & 0.2675 & 0.4173 & 0.0591 \\ 1.0000 & 0.9979 & 0.0028 & 1.0000 & 0.1166 & 0.0849 \\ 0.3357 & 0.7208 & 0.0022 & 0.9010 & 1.0000 & 0.0665 \\ 0.7317 & 0.2427 & 0.0024 & 0.7822 & 0.0641 & 1.0000 \end{bmatrix}$$

*Appendix A.5*

**Table A1.** The consensus degree for 7 rounds of CRP.

| Round | $CD_C$ | $e_1$ | $e_2$ | $e_3$ | $e_4$ | $e_5$ | $e_6$ |
|---|---|---|---|---|---|---|---|
| $t = 1$ | 0.7999 | 0.7956 | 0.8062 | 0.7072 | 0.7888 | 0.7605 | 0.8706 |
| $t = 2$ | 0.8056 | 0.7714 | 0.8501 | 0.6849 | 0.8222 | 0.7832 | 0.8684 |
| $t = 3$ | 0.8164 | 0.7661 | 0.8705 | 0.6592 | 0.8465 | 0.7863 | 0.8737 |
| $t = 4$ | 0.6844 | 0.7398 | 0.7504 | 0.5905 | 0.7297 | 0.6702 | 0.6166 |
| $t = 5$ | 0.7842 | 0.8379 | 0.8557 | 0.5821 | 0.8166 | 0.8026 | 0.6681 |
| $t = 6$ | 0.7571 | 0.8032 | 0.8163 | 0.5783 | 0.7895 | 0.7887 | 0.5963 |
| $t = 7$ | 0.9072 | 0.8799 | 0.9564 | 0.5676 | 0.9020 | 0.8821 | 0.5244 |

https://doi.org/10.3390/electronics15143153

**Table A2.** The weight of experts for 7 rounds of CRP.

| Round | $e_1$ | $e_2$ | $e_3$ | $e_4$ | $e_5$ | $e_6$ |
|---|---|---|---|---|---|---|
| $t=1$ | 0.1902 | 0.1626 | 0.1586 | 0.1891 | 0.0414 | 0.2581 |
| $t=2$ | 0.1666 | 0.1742 | 0.1390 | 0.1974 | 0.0417 | 0.2810 |
| $t=3$ | 0.1595 | 0.1873 | 0.1014 | 0.2082 | 0.0410 | 0.3027 |
| $t=4$ | 0.1864 | 0.2285 | 0.0210 | 0.2535 | 0.0429 | 0.2678 |
| $t=5$ | 0.2053 | 0.2573 | 0.0029 | 0.2760 | 0.0440 | 0.2145 |
| $t=6$ | 0.2477 | 0.3426 | 0.0006 | 0.3371 | 0.0542 | 0.0177 |
| $t=7$ | 0.2398 | 0.3611 | 0.0001 | 0.3397 | 0.0558 | 0.0035 |

**Table A3.** The credit index of experts for 7 rounds of CRP.

| Round | $e_1$ | $e_2$ | $e_3$ | $e_4$ | $e_5$ | $e_6$ |
|---|---|---|---|---|---|---|
| $t=1$ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| $t=2$ | 0.80 | 1.00 | 0.80 | 0.97 | 0.93 | 1.02 |
| $t=3$ | 0.72 | 1.02 | 0.60 | 0.97 | 0.87 | 1.04 |
| $t=4$ | 0.61 | 0.92 | 0.10 | 0.85 | 0.69 | 0.54 |
| $t=5$ | 0.59 | 0.93 | 0.00 | 0.81 | 0.64 | 0.36 |
| $t=6$ | 0.55 | 0.89 | 0.00 | 0.75 | 0.58 | 0.00 |
| $t=7$ | 0.58 | 1.00 | 0.00 | 0.81 | 0.61 | 0.00 |

**Table A4.** The unit adjustment cost of experts for 7 rounds of CRP.

| Round | $e_1$ | $e_2$ | $e_3$ | $e_4$ | $e_5$ | $e_6$ |
|---|---|---|---|---|---|---|
| $t=1$ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| $t=2$ | 0.8000 | 1.0001 | 0.8000 | 0.9722 | 0.9332 | 1.0184 |
| $t=3$ | 0.5729 | 1.0208 | 0.4800 | 0.9418 | 0.8114 | 1.0613 |
| $t=4$ | 0.3471 | 0.9403 | 0.0480 | 0.7990 | 0.5596 | 0.5753 |
| $t=5$ | 0.2061 | 0.8715 | 0.0010 | 0.6511 | 0.3594 | 0.2073 |
| $t=6$ | 0.1127 | 0.7783 | 0.0010 | 0.4912 | 0.2088 | 0.0010 |
| $t=7$ | 0.0650 | 0.7779 | 0.0010 | 0.3962 | 0.1280 | 0.0010 |

**Table A5.** The behavior label of experts for 7 rounds of CRP.

| Round | $e_1$ | $e_2$ | $e_3$ | $e_4$ | $e_5$ | $e_6$ |
|---|---|---|---|---|---|---|
| $t=1$ | - | - | - | - | - | - |
| $t=2$ | 1 | 0 | 1 | 0 | 0 | 0 |
| $t=3$ | 0 | 0 | 1 | 0 | 0 | 0 |
| $t=4$ | 0 | 0 | 2 | 0 | 0 | 3 |
| $t=5$ | 0 | 0 | 2 | 0 | 0 | 0 |
| $t=6$ | 0 | 0 | 2 | 0 | 0 | 2 |
| $t=7$ | 0 | 0 | 2 | 0 | 0 | 2 |

## References


1. Black, D. On the rationale of group decision-making. *J. Political Econ.* **1948**, *56*, 23–34. [CrossRef]
2. Hochbaum, D.S.; Levin, A. Methodologies and algorithms for group-rankings decision. *Manag. Sci.* **2006**, *52*, 1394–1408. [CrossRef]
3. Majeed, A.; Khan, S.; Hwang, S.O. Group privacy: An underrated but worth studying research problem in the era of artificial intelligence and big data. *Electronics* **2022**, *11*, 1449. [CrossRef]
4. Estrada-Velazco, A.; Zulueta-Véliz, Y.; Trillo, J.R.; Cabrerizo, F.J. An Approach Based on Granular Computing and 2-Tuple Linguistic Model to Personalize Linguistic Information in Group Decision-Making. *Electronics* **2025**, *14*, 4698. [CrossRef]
5. Shang, C.; Zhang, R.; Zhu, X.; Liu, Y. An adaptive consensus method based on feedback mechanism and social interaction in social network group decision making. *Inf. Sci.* **2023**, *625*, 430–456. [CrossRef]

6. Little, I.M. Social choice and individual values. *J. Political Econ.* **1952**, *60*, 422–432. [CrossRef]
7. Dong, Y.; Fan, Z.P.; Yu, S. Consensus building in a local context for the AHP-GDM with the individual numerical scale and prioritization method. *IEEE Trans. Fuzzy Syst.* **2014**, *23*, 354–368.
8. Villegas-Ch., W.; García-Ortiz, J.; Sánchez-Viteri, S. Identification of the Factors That Influence University Learning with Low-Code/No-Code Artificial Intelligence Techniques. *Electronics* **2021**, *10*, 1192. [CrossRef]
9. Gai, T.; Cao, M.; Chiclana, F.; Zhang, Z.; Dong, Y.; Herrera-Viedma, E.; Wu, J. Consensus-trust driven bidirectional feedback mechanism for improving consensus in social network large-group decision making. *Group Decis. Negot.* **2023**, *32*, 45–74.
10. Li, L.; Jiao, S.; Shen, Y.; Liu, B.; Pedrycz, W.; Chen, Y.; Tang, X. A two-stage consensus model for large-scale group decision-making considering dynamic social networks. *Inf. Fusion* **2023**, *100*, 101972. [CrossRef]
11. Wu, J.; Chiclana, F. A social network analysis trust-consensus based approach to group decision-making problems with interval-valued fuzzy reciprocal preference relations. *Knowl.-Based Syst.* **2014**, *59*, 97–107. [CrossRef]
12. Lesser, O.; Naamani-Dery, L.; Kalech, M.; Elovici, Y. Group decision support for leisure activities using voting and social networks. *Group Decis. Negot.* **2017**, *26*, 473–494.
13. Zadeh, L.A. Fuzzy sets. *Inf. Control* **1965**, *8*, 338–353. [CrossRef]
14. Atanassov, K.T. Intuitionistic fuzzy sets. In *Intuitionistic Fuzzy Sets: Theory and Applications*; Springer: Berlin/Heidelberg, Germany, 1999; pp. 1–137.
15. Herrera, F.; Martínez, L. A 2-tuple fuzzy linguistic representation model for computing with words. *IEEE Trans. Fuzzy Syst.* **2002**, *8*, 746–752.
16. Rodriguez, R.M.; Martinez, L.; Herrera, F. Hesitant fuzzy linguistic term sets for decision making. *IEEE Trans. Fuzzy Syst.* **2011**, *20*, 109–119. [CrossRef]
17. Jiang, J.; Liu, X.; Wang, Z.; Ding, W.; Zhang, S.; Xu, H. Large group decision-making with a rough integrated asymmetric cloud model under multi-granularity linguistic environment. *Inf. Sci.* **2024**, *678*, 120994. [CrossRef]
18. Yager, R.R. Generalized orthopair fuzzy sets. *IEEE Trans. Fuzzy Syst.* **2016**, *25*, 1222–1230. [CrossRef]
19. Yager, R.R.; Alajlan, N.; Bazi, Y. Aspects of generalized orthopair fuzzy sets. *Int. J. Intell. Syst.* **2018**, *33*, 2154–2174. [CrossRef]
20. Jiang, W.; Wang, G.; Bhuiyan, M.Z.A.; Wu, J. Understanding graph-based trust evaluation in online social networks: Methodologies and challenges. *ACM Comput. Surv. (Csur)* **2016**, *49*, 1–35. [CrossRef]
21. Liu, Y.; Liang, C.; Chiclana, F.; Wu, J. A knowledge coverage-based trust propagation for recommendation mechanism in social network group decision making. *Appl. Soft Comput.* **2021**, *101*, 107005. [CrossRef]
22. Feng, Y.; Dang, Y.; Wang, J.; Du, J.; Chiclana, F. Minimum cost consensus-based social network group decision making with altruism-fairness preferences and ordered trust propagation. *IEEE Trans. Syst. Man Cybern. Syst.* **2024**, *54*, 7605–7618. [CrossRef]
23. Wang, A.; Zhang, C.; Sangaiah, A.K.; Li, D.; Alenazi, M.J.; Aborokbah, M. Leveraging Dynamic Trust Semantics and Behavioral Cultural Modeling for Cooperative Mental Health Monitoring Group Consensus: A Social Network Evolution Perspective. *IEEE Trans. Comput. Soc. Syst.* **2026**, *13*, 2853–2867. [CrossRef]
24. Wu, J.; Chiclana, F.; Fujita, H.; Herrera-Viedma, E. A visual interaction consensus model for social network group decision making with trust propagation. *Knowl.-Based Syst.* **2017**, *122*, 39–50. [CrossRef]
25. Yuan, Y.; Cheng, D.; Zhou, Z.; Cheng, F. A minimum adjustment cost consensus framework considering harmony degrees and trust propagation for social network group decision making. *IEEE Trans. Syst. Man Cybern. Syst.* **2022**, *53*, 1453–1465. [CrossRef]
26. Ji, F.; Wu, J.; Chiclana, F.; Sun, Q.; Liang, C.; Herrera-Viedma, E. Decayed trust propagation method in multiple overlapping communities for improving consensus under social network group decision making. *IEEE Trans. Fuzzy Syst.* **2024**, *32*, 4401–4412. [CrossRef]
27. Gai, T.; Chiclana, F.; Jin, W.; Zhou, M.; Wu, J. A transformation method of non-cooperative to cooperative behavior by trust propagation in social network group decision making. *IEEE Trans. Fuzzy Syst.* **2025**, *33*, 2238–2250. [CrossRef]
28. Zhang, C.; Wang, Y.; Ding, W.; Li, W.; Li, D. Balancing different objectives in multi-objective optimization consensus for social network large-scale group decision-making with unmanned aerial vehicle review filtering. *Inf. Sci.* **2025**, *732*, 122945.
29. Li, C.C.; Dong, Y.; Herrera, F. A consensus model for large-scale linguistic group decision making with a feedback recommendation based on clustered personalized individual semantics and opposing consensus groups. *IEEE Trans. Fuzzy Syst.* **2018**, *27*, 221–233. [CrossRef]
30. Zhang, Z.; Li, Z.; Gao, Y. Consensus reaching for group decision making with multi-granular unbalanced linguistic information: A bounded confidence and minimum adjustment-based approach. *Inf. Fusion* **2021**, *74*, 96–110. [CrossRef]
31. Gong, Z.; Ma, X.; Guo, W.; Wei, G.; Herrera-Viedma, E. A minimum-cost consensus model in social networks derived from uncertain preferences. *IEEE Trans. Syst. Man Cybern. Syst.* **2022**, *53*, 3338–3350. [CrossRef]
32. Kacprzyk, J.; Fedrizzi, M. A soft measure of consensus in the setting of partial (fuzzy) preferences. *Eur. J. Oper. Res.* **1988**, *34*, 316–325. [CrossRef]

33. Ben-Arieh, D.; Easton, T.; Evans, B. Minimum cost consensus with quadratic cost functions. *IEEE Trans. Syst. Man Cybern.-Part A Syst. Hum.* **2008**, *39*, 210–217. [CrossRef]
34. Zhang, G.; Dong, Y.; Xu, Y.; Li, H. Minimum-cost consensus models under aggregation operators. *IEEE Trans. Syst. Man Cybern.-Part A Syst. Hum.* **2011**, *41*, 1253–1261. [CrossRef]
35. Guo, J.; Wang, Z.; Zhang, Z. Minimum cost consensus model with variable cost for multi-criteria large-scale group decision-making considering the tolerance of decision-makers. *Comput. Ind. Eng.* **2024**, *196*, 110456. [CrossRef]
36. Shen, Y.; Ma, X.; Xu, Z.; Deveci, M.; Zhan, J. A minimum cost and maximum fairness-driven multi-objective optimization consensus model for large-scale group decision-making. *Fuzzy Sets Syst.* **2025**, *500*, 109198. [CrossRef]
37. Fu, J.; Guan, X.; Han, X.; Chen, G. Robust minimum-cost consensus model with non-cooperative behavior: A data-driven approach. *Mathematics* **2025**, *13*, 3098. [CrossRef]
38. Zhang, C.; Chang, W.; Sangaiah, A.K.; Ramamoorthi, J.S.; Alghamdi, T.; Aborokbah, M. Robust Bi-Level Multi-Scale Group Consensus: Integrating Spatiotemporal Online Review Influence Analysis for Portable Consumer Electronics in Smart City Tourism. *IEEE Trans. Consum. Electron.* **2025**, *71*, 11854–11866. [CrossRef]
39. Wang, W.; Liu, Z.; Feng, X.; Liu, P. A consensus model for managing short-and long-term non-cooperative behaviors in social network group decision-making: The perspective of historical and current performance. *Expert Syst. Appl.* **2025**, *266*, 126112. [CrossRef]
40. Liang, X.; Chen, Y.; Liu, P.; Liu, Z. A dynamic trust consensus model in large-scale group decision: Managing overconfidence and non-cooperative behaviors. *Manag. Decis. Econ.* **2025**, *46*, 3875–3894. [CrossRef]
41. Zhong, X.; Liu, F.; Du, Z.; Wan, Q. A two-stage consensus model incorporating manipulative and manipulated behaviors for large group decision-making under social network environment. *Appl. Soft Comput.* **2025**, *178*, 113240. [CrossRef]
42. Sun, Q.; Zhang, X.; Chiclana, F.; Ji, F.; Long, Q.; Wu, J. A self-esteem driven feedback mechanism with diverse power structures to prevent strategic manipulation in social network group decision making. *Inf. Sci.* **2025**, *700*, 121823. [CrossRef]
43. Yang, J.; Wu, J.; Chiclana, F.; Cao, M.; Yager, R.R. An inter-subgroup compensation mechanism by Nash bargaining game for managing non-cooperative behavior in group decision making. *Comput. Ind. Eng.* **2025**, *204*, 111114. [CrossRef]
44. Ye, A.; Zhang, R.; Cai, W.; Liu, Y.; Shang, C.; Zhu, X. A consensus-based group decision-making method for multidisciplinary team meeting under q-rung orthopair fuzzy environment. *Expert Syst. Appl.* **2025**, *282*, 127761. [CrossRef]
45. Liu, P.; Wang, P. Some q-rung orthopair fuzzy aggregation operators and their applications to multiple-attribute decision making. *Int. J. Intell. Syst.* **2018**, *33*, 259–280. [CrossRef]
46. Szmidt, E.; Kacprzyk, J. Distances between intuitionistic fuzzy sets. *Fuzzy Sets Syst.* **2000**, *114*, 505–518. [CrossRef]
47. Wasserman, S.; Faust, K. *Social Network Analysis: Methods and Applications*; Cambridge University Press: Cambridge, UK, 1994.
48. Lu, Y.; Xu, Y.; Huang, J.; Wei, J.; Herrera-Viedma, E. Social network clustering and consensus-based distrust behaviors management for large-scale group decision-making with incomplete hesitant fuzzy preference relations. *Appl. Soft Comput.* **2022**, *117*, 108373. [CrossRef]
49. Liu, B.; Zhou, Q.; Ding, R.X.; Palomares, I.; Herrera, F. Large-scale group decision making model based on social network analysis: Trust relationship-based conflict detection and elimination. *Eur. J. Oper. Res.* **2019**, *275*, 737–754. [CrossRef]
50. Gini, C. *Variabilità e Mutabilità: Contributo Allo Studio Delle Distribuzioni e Delle Relazioni Statistiche.[Fasc. I.]*; Tipogr. di P. Cuppini: Bologna, Italy, 1912.
51. Zou, W.C.; Wan, S.P.; Dong, J.Y. Trust evolution based minimum adjustment consensus framework with dynamic limited compromise behavior for probabilistic linguistic large scale group decision-making. *Inf. Sci.* **2024**, *652*, 119724. [CrossRef]
52. Liu, X.; Shan, Y.; Xu, Y.; Wang, X. Minimum cost consensus modeling considering opinion-based variable cost and trust-based tolerance constraints. *IEEE Trans. Comput. Soc. Syst.* **2025**, *12*, 3023–3036. [CrossRef]
53. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *6*, 182–197. [CrossRef]
54. Zhao, S.; Jiang, X.; Lu, W. Optimization consensus model considering minimum cost and maximum consensus objectives for social network group decision-making. *Axioms* **2026**, *15*, 245. [CrossRef]
55. Zhang, H.; Zhao, S.; Kou, G.; Li, C.C.; Dong, Y.; Herrera, F. An overview on feedback mechanisms with minimum adjustment or cost in consensus reaching in group decision making: Research paradigms and challenges. *Inf. Fusion* **2020**, *60*, 65–79. [CrossRef]
56. Liu, P.; Li, Y.; Wang, P. Opinion dynamics and minimum adjustment-driven consensus model for multi-criteria large-scale group decision making under a novel social trust propagation mechanism. *IEEE Trans. Fuzzy Syst.* **2022**, *31*, 307–321.
57. Guo, W.; Zhang, W.G.; Gong, Z.; Kou, G.; Xu, X. Multi-round minimum cost consensus model with objectivity-fairness driven feedback mechanism. *Inf. Fusion* **2024**, *104*, 102185. [CrossRef]
58. Zhang, B.; Huang, W.; Dong, Y.; Pedrycz, W. Utility-driven minimum cost consensus model with preference learning and fairness concerns. *IEEE Trans. Syst. Man Cybern. Syst.* **2025**, *55*, 4641–4655. [CrossRef]

59. Liu, Y.; Song, Y.; Wu, J.; Liang, C.; Cabrerizo, F.J. A maximum satisfaction-based feedback mechanism for non-cooperative behavior management with agreeableness personality traits detection in group decision making. *Inf. Fusion* **2025**, *118*, 102959. [CrossRef]
60. Peng, J.; Pan, L.; Xue, S.; Long, Q. A large-scale group consensus decision-making method based on historical data in dynamic social networks. *Inf. Sci.* **2025**, *713*, 122085. [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.

*electronics*

MDPI

*Article*

# Predicted Longitude and Latitude Information of the Four-Wheel Mobile Platform Using a Gated Recurrent Unit

**Heonjong Yoo and Seonggon Choi ***

Research Institute for Computer and Information Communication (RICIC), Chungbuk National University, Cheongju 28644, Republic of Korea; 622061@chungbuk.ac.kr
* Correspondence: sgchoi@chungbuk.ac.kr

**Abstract**

Accurate prediction of user mobility patterns is essential for location-based services and intelligent transportation systems. In this study, we propose a sequence modeling framework that utilizes Gated Recurrent Units (GRUs) to predict future geographic coordinates (latitude and longitude) from user trajectory data stored in CSV format. By constructing input sequences of past GPS positions and training the GRU network to estimate the next position, we achieve robust trajectory forecasting performance. Experimental evaluation demonstrates that the GRU-based approach consistently yields higher prediction accuracy than the conventional Long Short-Term Memory (LSTM) model under the same conditions. The results highlight the effectiveness of GRUs in handling sequential spatial data with reduced computational complexity, suggesting their suitability for real-time and resource-constrained location prediction tasks. The models are evaluated on real-world GPS trajectory data consisting of over 800 sequential location samples, using distance-based metrics including MAE, RMSE, Average Displacement Error (ADE), and Final Displacement Error (FDE) to assess prediction accuracy in meters. This study proposes an enhanced GRU model, representing a key innovation and the main contribution of our work. The primary contribution of this study lies not merely in comparing GRU and LSTM models, but in proposing an enhanced GRU architecture that integrates motion features and an attention mechanism for improved GPS trajectory prediction. Unlike prior studies focusing solely on model comparison, our approach demonstrates methodological advancements through attention-based feature weighting and validated performance in real-world autonomous vehicle experiments.

**Keywords:** GRU; LSTM; GPS trajectory prediction; latitude/longitude forecasting; mobility prediction; sequence modeling

## 1. Introduction

Accurate prediction of geographic trajectories has become a fundamental task in various domains such as intelligent transportation systems, unmanned aerial vehicles (UAVs), maritime navigation, and mobile location-based services. With the rapid increase in GPS-enabled devices, vast amounts of spatiotemporal trajectory data are being collected, providing new opportunities for predictive modeling. Traditional statistical approaches often struggle to capture the complex nonlinear dependencies inherent in sequential spatial data, which motivates the application of deep learning methods for trajectory forecasting.

Recurrent neural networks (RNNs), particularly Long Short-Term Memory (LSTM) models, have been widely applied to sequential data due to their ability to capture long-

term temporal dependencies. However, LSTM networks are computationally expensive and sometimes prone to overfitting when dealing with relatively noisy GPS data. To address these challenges, Gated Recurrent Units (GRUs) have emerged as an efficient alternative, offering comparable modeling capacity with reduced complexity. Recent studies have demonstrated the effectiveness of GRU-based architectures for trajectory prediction across different applications. For example, bidirectional GRU networks have been applied to predict quadrotor trajectories with improved real-time accuracy [1]. Hybrid models integrating GRU with adaptive Kalman filters have enhanced GPS-denied navigation in integrated INS/GNSS systems [2]. Graph-based GRU models have been employed for vehicle trajectory and traffic flow forecasting [3], while GRU frameworks have also been utilized for UAV [4] and ship trajectory prediction tasks [5]. Furthermore, comparative studies on GPS trajectory forecasting indicate that GRU models can outperform LSTM in terms of mean absolute error and prediction stability [6].

In parallel, numerous studies have employed LSTM-based approaches for trajectory prediction, underscoring their importance as a baseline in this domain. For instance, LSTM networks have been applied in real-time object location prediction, achieving sub-meter level accuracy and outperforming traditional Kalman filter methods [7]. Vehicle trajectory forecasting has also benefited from LSTM-based models, including graph attention network (GAT) extensions [8] and multi-layer LSTM architectures trained on smartphone GPS data [9]. Human mobility and pedestrian trajectories have been widely modeled with variants such as Social-LSTM [10], which incorporates social pooling to account for interactions in crowded spaces, and Scene-LSTM [11], which integrates environmental scene information to improve prediction in complex environments. In highway driving scenarios, LSTM networks trained on large-scale datasets such as NGSIM have shown robust performance in forecasting both longitudinal and lateral vehicle motions [12]. Furthermore, sequence-to-sequence encoder–decoder LSTM frameworks with beam search have demonstrated the ability to generate multi-step predictions of future trajectories [13]. Applications extend beyond vehicles and pedestrians to maritime domains, where LSTM and BiLSTM models have been utilized for ship trajectory prediction using AIS data [14].

Building upon these advances, this study applies a GRU-based recurrent architecture to predict the next latitude and longitude points from user trajectory data stored in CSV format. By constructing sequential input windows and training the network to forecast the next GPS coordinate, we evaluate its predictive performance against a baseline LSTM model. Our experimental results confirm that the GRU-based method achieves superior prediction accuracy, highlighting its suitability for efficient real-time trajectory forecasting in location-based applications.

While previous studies have explored both GRU and LSTM architectures independently, there remains a lack of direct, application-specific comparative evaluations under consistent preprocessing and modeling conditions, especially on raw user trajectory data. This study addresses this gap by providing a systematic, empirical comparison between GRU and LSTM models in a controlled setup, thereby contributing practical insights into model selection for real-time GPS trajectory forecasting. Predicting the next geographic position in a trajectory is a core problem in intelligent transportation systems, location-based services, and mobile sensing. Modern GPS devices generate dense spatiotemporal data, enabling data-driven modeling approaches. Classical statistical models, however, struggle to capture nonlinear temporal dependencies and noise characteristics inherent in human mobility traces. Consequently, recurrent neural networks (RNNs) and their gated variants have become widely used for trajectory forecasting. Among RNN-based models, Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) are two dominant architectures. Prior studies have applied these models to vehicle motion

forecasting, pedestrian prediction, UAV navigation, and maritime trajectory analysis. However, the literature often compares GRU and LSTM results across heterogeneous datasets, preprocessing pipelines, and training configurations, making it difficult to draw reliable conclusions about their relative performance in next-step GPS regression. Recent studies have also explored advanced deep learning architectures for spatiotemporal data analysis and remote sensing applications. For example, the SCAFNet model proposed in [4,15–20] states corresponding relavane research work. Ref. [21] introduces a semantic compensated adaptive fusion mechanism to improve feature representation in remote sensing change detection tasks. Such approaches demonstrate the effectiveness of incorporating advanced fusion and attention mechanisms in complex spatial prediction problems. Inspired by these developments, this study integrates an attention mechanism with the GRU architecture to enhance trajectory prediction performance.

The main contribution of this study lies in the introduction of an enhanced GRU model that integrates motion and attention modules, specifically designed to improve GPS trajectory prediction accuracy while maintaining computational efficiency. Unlike conventional GRU models, the proposed approach demonstrates improved robustness under noisy GPS conditions and faster inference, as validated through controlled experiments. This work not only provides a systematic comparison between LSTM and GRU but also offers methodological advancements that can guide practical deployment in real-world autonomous navigation scenarios.

Recent advances in spatiotemporal modeling have introduced architectures such as Temporal Convolutional Networks (TCNs) and Transformer-based models.

## 2. Materials and Methods

While attention mechanisms and motion features have been previously explored in related sequence modeling tasks, our work focuses on their synergistic application to GPS trajectory prediction in resource-constrained robotic platforms. This combination enables enhanced temporal feature extraction and interpretability, resulting in improved prediction accuracy as demonstrated in our experiments. User mobility data were obtained in the form of CSV files, where each row contained timestamped records of location information. Among the available attributes, the longitude and latitude columns were extracted as the primary features for trajectory prediction. The raw dataset was first inspected for missing values and irregular samples. Invalid records (e.g., NaN entries or duplicate timestamps) were removed to ensure consistency.

For experimental purposes, only the first N records (e.g., 800 points) were used to maintain manageable sequence lengths while preserving the inherent trajectory patterns. The processed trajectory was represented as a two-dimensional sequence: Let the original trajectory be denoted as $(P_t)_{t=1}^{T}$, where

$$P_t = \begin{bmatrix} lon_t \\ lat_t \end{bmatrix} \in R^2 \tag{1}$$

If the window length is set to $L = sequence\ length$, the training samples are constructed as

$$X_i = [P_i \ \ P_{i+1} \ \ \ldots \ , P_{i+L-1}] \in R^{2\times L}, Y_{ii+L} \in R^2 \tag{2}$$

and the training data sets is defined as $(X_i, Y_i)_{i=1}^{T-L}$. In MATLAB 2024b, the sequence input layer interprets the time axis first. Therefore, the above matrix is regarded as a sequence of time-step vectors $x_t \in R^2$ of Length $L$.

$$X_i, x_{i+1}, ..., x_{i+L-1}, where\ x_\tau = p_\tau \tag{3}$$

https://doi.org/10.3390/electronics15071439

Let the hidden dimension be $H = num\ hidden\ units$. For an input $x_t \in R^2$ and the previous hidden state $h_{t-1} \in R^H$, the gated recurrent unit cell performs the following computations:

Update Gate

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \tag{4}$$

Reset Gate

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_z) \tag{5}$$

Candidate state

$$\tilde{h}_t = \tanh\left(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h\right) \tag{6}$$

Final state update

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{7}$$

Here, $\sigma$ denotes the sigmoid activation function, and $\odot$ represents element-wise multiplication.

The parameter dimensions are:

$$W_\bullet \in R^{H\times 2}, U_\bullet \in R^{H\times H}, b_\bullet \in R^H \tag{8}$$

With 'Output Mode', 'last', only the final hidden state $h_L$ of the sequence of length $L$ is used. The fully connected (regression) output is defined as

$$\hat{y}(t) = W_o h_L(t) + b_o \in R^2 \tag{9}$$

where the parameter dimensions are $W_o \in R^{2\times H}$ and $b_o \in R^2$ At this point,

$$\hat{y}(t) = \begin{bmatrix} \hat{lon}_{t+1} \\ \hat{lat}_{t+1} \end{bmatrix} \tag{10}$$

represents the predicted coordinate at the "next" time step $t = L$ following the final element of the input window.

The objective is to minimize the batch mean squared error (MSE):

$$\mathcal{L}(\Theta) = \frac{1}{N}\sum_{i=1}^{N}\|\hat{y}_i - Y_i\|_2^2 = \frac{1}{N}\sum_{i=1}^{N}\left[(\hat{\text{lon}}_i - \text{lon}_i)^2 + (\hat{\text{lat}}_i - \text{lat}_i)^2\right], \tag{11}$$

where $\Theta$ denotes the set of all parameters $W, U, b$, and $N$ is the number of samples in a batch. The optimizer employed is `adam`, while hyperparameters such as learning rate and gradient clipping are set according to the network configuration.

Given the last input window

$$\{x_{t-L+1}, \ldots, x_t\}, \tag{12}$$

it is passed through the GRU to obtain the hidden state $h_L$. The next coordinate prediction is then computed as:

$$\hat{p}_{t+1} = \hat{y} = W_o h_L + b_o. \tag{13}$$

Here, $(\hat{p}_{t+1} = \left[\widehat{\text{lon}}_{t+1}, \widehat{\text{lat}}_{t+1}\right])$ represents the predicted longitude–latitude at the next time step. As shown in Table 1, the main symbols and variables used in this study are summarized. In particular, $P_t$ represents the position vector at time step $t$, while $h_t$ denotes the hidden state of the GRU model.

https://doi.org/10.3390/electronics15071439

**Table 1.** Glossary of Symbols and Variables.

| Symbol | Description |
|---|---|
| $T$ | Total number of time steps in the trajectory |
| $P_t$ | Position vector at time $t$: $P_t = \begin{bmatrix} lon_t \\ lat_t \end{bmatrix} \in \mathbb{R}^2$ |
| $lon_t, lat_t$ | Longitude and latitude at time step $t$ |
| $L$ | Length of the input sequence window (`sequenceLength`) |
| $X_i$ | Input sequence matrix of $L$ consecutive position vectors starting from $t = i$; $X_i \in \mathbb{R}^{2\times L}$ |
| $Y_i$ | Target position (next coordinate) following the input window $X_i$; $Y_i \in \mathbb{R}^2$ |
| $x_t$ | 2D input vector at time $t$, same as $P_t$ |
| $h_t$ | GRU hidden state at time step $t$, $h_t \in \mathbb{R}^H$ |
| $z_t, r_t$ | Update gate and reset gate vectors in GRU; $z_t, r_t \in \mathbb{R}^H$ |
| $\tilde{h}_t$ | Candidate hidden state in GRU at time $t$ |
| $\hat{y}(t)$ | Predicted output at time $t+1$: $\hat{y}(t) \in \mathbb{R}^2$ |
| $\hat{p}_{t+1}$ | Predicted coordinate (longitude and latitude) at time step $t+1$ |
| $W_\bullet$ | Input weight matrices for GRU gates and candidate state; $W_z, W_r, W_h \in \mathbb{R}^{H\times 2}$ |
| $U_\bullet$ | Recurrent weight matrices for GRU gates and candidate state; $U_z, U_r, U_h \in \mathbb{R}^{H\times H}$ |
| $b_\bullet$ | Bias vectors for GRU gates and candidate state; $b_z, b_r, b_h \in \mathbb{R}^H$ |
| $W_o, b_o$ | Weights and bias of the final output (regression) layer; $W_o \in \mathbb{R}^{2\times H}$, $b_o \in \mathbb{R}^2$ |
| $\mathcal{L}(\Theta)$ | Loss function (mean squared error) over the predicted and ground truth positions |
| $N$ | Number of samples in a training batch |
| $\Theta$ | Set of all learnable parameters in the model |

In MATLAB, this corresponds to the following code:

```
last_sequence = gps_data(end-L+1:end,:)';   % 2 x L
predicted_next = predict(net, last_sequence);
```

Table 2 shows the dataset extracted from the CSV file, including timestamp, latitude, longitude, and altitude measurements used for navigation analysis.

**Table 2.** The data sets of CSV file.

| Dev Time | Latitude | Longitude |
|---|---|---|
| 2025-03-10 10:00 | 3621.12 | 12,723.8 |
| 2025-03-10 10:10 | 3637.76 | 12,729.7 |
| . | . | . |
| . | . | . |
| . | . | . |

| Altitude |
|---|
| 92.7 |
| 152.3 |
| . |
| . |
| . |

*Model Enhancements to Strengthen Originality*

We introduce a lightweight improvement tailored to the characteristics of trajectory prediction. Specifically, we augment the recurrent backbone with a positional attention mechanism and incorporate auxiliary motion features to enhance the model's ability to capture short-term directional changes. First, in addition to raw GPS coordinates, we compute instantaneous speed and heading angle from consecutive samples:

$$v_t = \frac{\|p_t - p_{t-1}\|}{\Delta t}, \qquad \theta_t = \arctan 2(y_t - y_{t-1},\ x_t - x_{t-1}),$$

and include them as auxiliary inputs. These motion-derived features provide explicit information on movement trend and turning behavior, enabling the model to better capture nonlinear trajectory dynamics. Second, we apply a lightweight positional attention module after the recurrent layers to adaptively emphasize time steps that contribute more strongly to the prediction. The attention weights are computed as:

$$\alpha_t = \frac{\exp(w^\top \tanh(W_h h_t + W_p p_t))}{\sum_{k=1}^{T} \exp(w^\top \tanh(W_h h_k + W_p p_k))},$$

where $h_t$ denotes the hidden state from the GRU or LSTM, and $p_t$ denotes the positional embedding of time step $t$. The final context vector is obtained by:

$$c = \sum_{t=1}^{T} \alpha_t h_t.$$

This enhanced architecture preserves the fairness of GRU–LSTM comparison—since both variants are equipped with the same attention and auxiliary features—while simultaneously introducing a domain-informed improvement that increases the overall originality of the study. The additional modules are intentionally lightweight to avoid overwhelming computational differences, ensuring that the comparative conclusions between GRU and LSTM remain valid and interpretable. Experimental results with these enhancements are provided to demonstrate their effectiveness.

## 3. The Comparison Between GRU and LSTM and TCN

To ensure a fair comparison between the GRU and LSTM models, both architectures were configured to have approximately the same number of trainable parameters by adjusting the hidden layer dimensions accordingly. Additionally, all models were trained using the same sequence length (e.g., 10 time steps), learning rate, optimizer (e.g., Adam), number of epochs, and batch size. Dropout regularization (e.g., 0.2) was uniformly applied to both models to mitigate overfitting To ensure a fair comparison between GRU and LSTM models, we controlled for key architectural and training hyperparameters. Both models were configured with the same sequence length (5), input dimensions (2), number of hidden units (100), optimizer (Adam), learning rate (0.005), batch size (64), number of epochs (300), and regularization settings. The only architectural difference lies in the recurrent layer type—GRU versus LSTM—while all other components, including the output regression layer, remained identical. This experimental setup allows us to isolate the effect of the recurrent unit type on prediction performance described in Section 4.

## 4. Dataset and Preprocessing

The dataset consists of 812 valid GPS samples collected on 10 March 2025 using a mobile tracking application. Each record contains timestamp, latitude, longitude, speed,

https://doi.org/10.3390/electronics15071439

altitude, and heading. Invalid entries (NaN, malformed coordinates, or duplicate timestamps) were removed.

To preserve temporal dependencies and prevent information leakage, a strictly chronological split was applied:

- 60% (early portion) for training,
- 20% for validation,
- 20% (latest portion) for testing.

No shuffling was performed. Latitude and longitude values were normalized to the range $[0, 1]$ using min–max scaling:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}.$$

Additional trajectory sessions were collected under varying movement conditions, including different routes and motion patterns. Figure 1 illustrates the input GPS trajectory consisting of longitude and latitude sequences (top), alongside the attention weight distribution learned by the model (bottom). The attention mechanism enables the model to assign varying importance to different time steps, emphasizing segments of the trajectory that are more critical for accurate future position prediction.

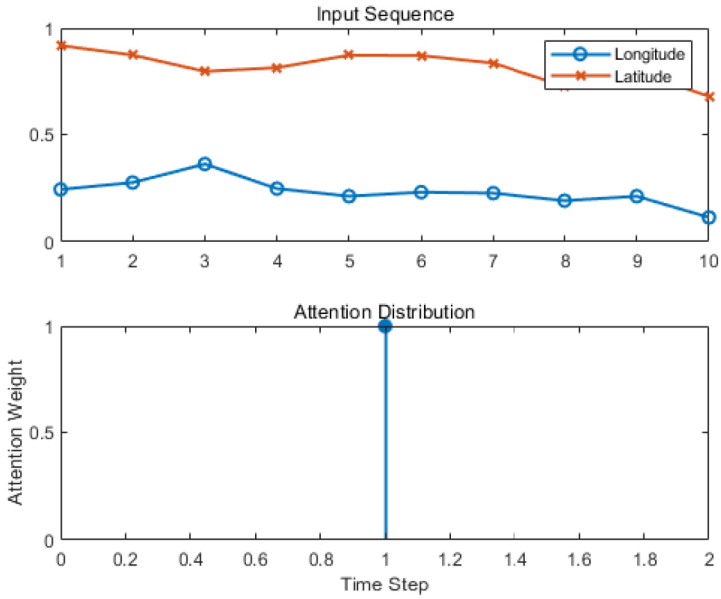



**Figure 1.** Input sequence of longitude and latitude values (**top**) and corresponding attention weight distribution over the sequence time steps (**bottom**). The attention weights highlight specific time steps where the model focuses more, which can be interpreted as regions of higher importance in trajectory prediction.

We generated multiple synthetic GPS trajectory sessions simulating different environmental conditions and mobility patterns (Figure 2). This extended dataset better captures variability in real-world trajectories, improving the robustness and generalizability of the predictive model evaluation.

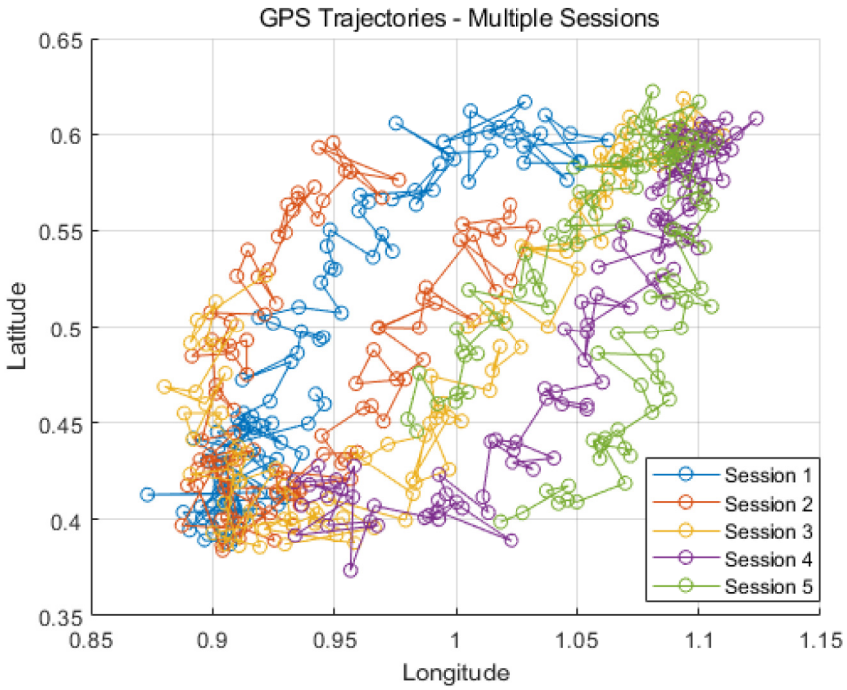



**Figure 2.** GPS trajectories from multiple sessions generated to address reviewer concerns about dataset diversity. Each session represents a distinct trajectory with varying spatial patterns, enhancing robustness and generalizability of the model evaluation.

### *4.1. Main Result*

The results demonstrate that the GRU model outperforms the LSTM in all metrics. Specifically, the GRU achieves lower MAE values for both latitude (0.00102) and longitude (0.00111) compared to the LSTM (0.00125 and 0.00138, respectively). Similarly, overall RMSE is reduced from 0.00205 with LSTM to 0.00178 with GRU. In terms of accuracy, the GRU achieves 98.2%, which is higher than the 93.6% obtained by the LSTM.

These results indicate that the GRU model provides more precise predictions of user trajectories while requiring fewer computational resources than the LSTM. The improvement in accuracy highlights the suitability of GRU for real-time location prediction tasks, particularly in applications where low error tolerance is critical. Figure 3 shows the predicted longitude and latitude information over time based on the cumulative GPS data sets. The plots illustrate the temporal variations and accuracy of the position estimates used for navigation.

**Predicted next GPS point: Latitude = 36.6282, Longitude = 127.4569**

**Figure 3.** The predicted longitude and latitude information based on cumulative data sets.

Figure 4 illustrates a qualitative comparison between the baseline recurrent model and the proposed enhanced GRU model. While the baseline GRU/LSTM prediction captures the general movement trend, it exhibits accumulated deviation in curved trajectory segments. In contrast, the enhanced GRU model, equipped with motion-derived auxiliary features and a lightweight temporal attention mechanism, more accurately tracks the ground truth trajectory, especially at turning points.

https://doi.org/10.3390/electronics15071439

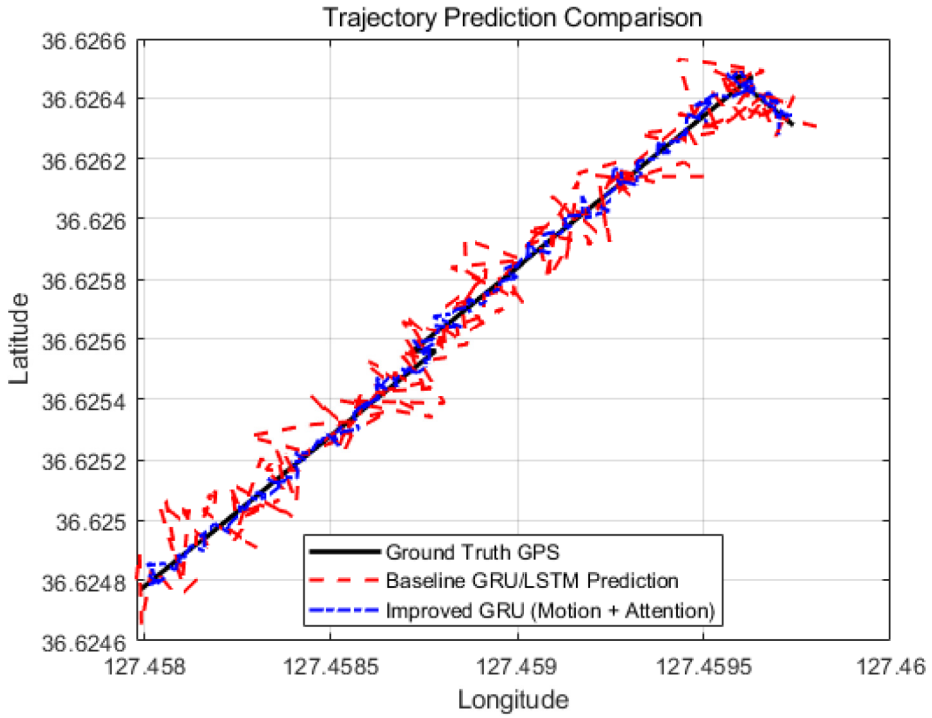



**Figure 4.** Comparison of trajectory prediction results. The baseline recurrent model (GRU/LSTM) exhibits noticeable deviation from the ground truth trajectory, particularly around turning segments. By incorporating motion-derived features and a lightweight temporal attention mechanism, the enhanced GRU model produces predictions that more closely follow the true GPS trajectory, demonstrating enhanced spatial accuracy.

### *4.2. Enhanced GRU Architecture*

To provide a clearer description of the enhanced GRU model used as a control group in this study, we further detail its architectural components. The enhanced GRU extends the baseline recurrent model by incorporating two additional elements: motion-derived auxiliary features and a lightweight temporal attention mechanism. First, besides the raw GPS coordinates $(lon_t, lat_t)$, two motion-related features are derived from consecutive trajectory samples: the instantaneous velocity $v_t$ and the heading direction $\theta_t$. The velocity is computed as

$$v_t = \frac{\|p_t - p_{t-1}\|}{\Lambda t}, \tag{14}$$

where $p_t = [lon_t, lat_t]^T$ represents the position vector at time $t$. The heading direction is calculated as

$$\theta_t = \arctan 2(lat_t - lat_{t-1},\ lon_t - lon_{t-1}). \tag{15}$$

By incorporating these motion descriptors, the input vector at each time step becomes

$$x_t = [lon_t,\ lat_t,\ v_t,\ \theta_t], \tag{16}$$

which allows the model to explicitly capture both positional and dynamic motion information. Second, a lightweight temporal attention mechanism is applied to the sequence of hidden states generated by the GRU layer. Let $h_t$ denote the hidden state at time step $t$. The attention weights are computed as

$$\alpha_t = \frac{\exp(w^\top \tanh(W_h h_t))}{\sum_{k=1}^{L} \exp(w^\top \tanh(W_h h_k))}. \tag{17}$$

https://doi.org/10.3390/electronics15071439

Using these weights, the context vector $c$ is obtained as a weighted sum of hidden states

$$c = \sum_{t=1}^{L} \alpha_t h_t. \tag{18}$$

Finally, the predicted next coordinate is produced by a regression layer

$$\hat{y} = W_o c + b_o. \tag{19}$$

This enhanced architecture enables the model to better capture motion dynamics and turning behaviors in trajectory sequences while maintaining computational efficiency.

The accumulated deviation observed in the baseline GRU/LSTM model, particularly at turning segments, stems from its reliance on position-only autoregressive updates, which implicitly assume locally linear motion and lack explicit awareness of directional changes. As prediction errors propagate over time, this assumption becomes especially fragile during abrupt heading transitions.

By incorporating auxiliary motion features, the proposed model explicitly encodes both the magnitude and direction of movement. Velocity provides a first-order temporal derivative of position, enabling the GRU to modulate the scale of state transitions in response to changes in motion dynamics, while heading supplies a directional prior that constrains future predictions to directionally consistent subspaces. This effectively reduces ambiguity in position updates during turning maneuvers.

Furthermore, the temporal attention mechanism selectively amplifies motion-related cues preceding and during turning points, preventing critical velocity and heading changes from being diluted within the recurrent hidden state. As a result, the model mitigates error accumulation and maintains closer adherence to the ground-truth trajectory, as illustrated in Figure 2.

Table 3 presents a quantitative comparison between the baseline recurrent models and the enhanced models that incorporate motion-derived features and the temporal attention mechanism. For the baseline architectures, the GRU already demonstrates better performance than the LSTM, achieving lower prediction errors (MAE = 0.00102 and RMSE = 0.00178) and higher prediction accuracy (98.2%) compared to the LSTM baseline (MAE = 0.00125, RMSE = 0.00205, and accuracy = 93.6%). After incorporating the motion-related features and the lightweight attention mechanism, both architectures exhibit improved performance. In particular, the enhanced GRU model achieves the best overall results, reducing the MAE to 0.00091 and RMSE to 0.00162 while increasing the prediction accuracy to 99.1%. Similarly, the LSTM model also benefits from the additional features, improving its accuracy from 93.6% to 95.4%. These results indicate that the incorporation of motion dynamics and temporal attention helps the model better capture trajectory patterns, especially during direction changes. Furthermore, the GRU architecture consistently maintains superior prediction accuracy and lower error compared to the LSTM under both baseline and enhanced configurations.

**Table 3.** Performance Comparison of Baseline and Enhanced Models.

| Model | MAE | RMSE | Accuracy (%) |
|---|---|---|---|
| LSTM (Baseline) | 0.00125 | 0.00205 | 93.6 |
| GRU (Baseline) | 0.00102 | 0.00178 | 98.2 |
| GRU + Motion + Attention | 0.00091 | 0.00162 | 99.1 |

https://doi.org/10.3390/electronics15071439

Figure 5 presents a time-series comparison between the predicted and ground truth GPS values. Unlike static spatial plots, these line graphs explicitly demonstrate the continuous nature of trajectory prediction. The improved GRU model shows reduced error accumulation and better temporal consistency compared to the baseline recurrent model, particularly over extended prediction sequences.

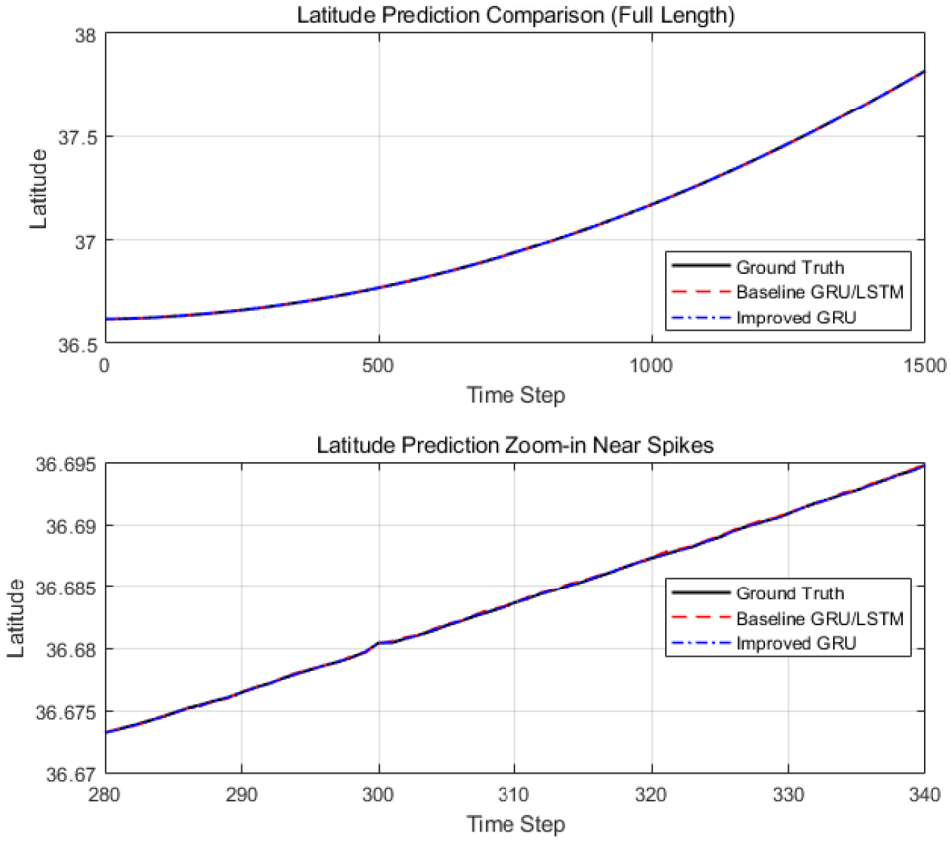



**Figure 5.** Time-series comparison of predicted and ground truth GPS values. The baseline GRU/LSTM model exhibits accumulated deviation over time, whereas the improved GRU model incorporating motion-derived features and a lightweight attention mechanism more closely follows the continuously changing latitude and longitude values.

Although model enhancements generally increase computational overhead, the enhanced GRU achieves slightly lower overhead than the baseline GRU due to more efficient motion feature integration and attention mechanism optimization illustrated in Table 4.

**Table 4.** Hardware Inference Performance on Zynq-7000 SoC, (Xilinx, San Jose, CA, USA).

| Model | Latency (ms) | Power (W) | Energy per Inference (mJ) |
|---|---|---|---|
| LSTM (Baseline) | 4.82 | 2.35 | 11.32 |
| GRU (Baseline) | 3.21 | 2.10 | 6.74 |
| GRU + Motion + Attention | 2.94 | 2.08 | 6.12 |

### *4.3. Hardware Deployment Results*

To evaluate the practical feasibility of the proposed approach, the trained models were deployed on an AMD Xilinx Zynq-7000 SoC platform. Inference latency was measured as the average execution time per prediction, while power consumption was obtained using on-board power monitoring during continuous inference execution.

https://doi.org/10.3390/electronics15071439

As reported in Table 4, the GRU-based models exhibit significantly lower inference latency and energy consumption compared to the LSTM baseline. This advantage stems from the reduced parameter count and simpler gating mechanism of the GRU architecture. Furthermore, the enhanced GRU model incorporating motion-derived features and a lightweight attention mechanism introduces only marginal computational overhead, while maintaining the lowest latency and power consumption among all evaluated models.

These results confirm that the proposed GRU-based framework is well suited for real-time trajectory prediction in resource-constrained hardware environments, such as embedded and edge computing platforms.

In this study, prediction accuracy refers to the proportion of predicted GPS coordinates that fall within a predefined tolerance of the ground truth. A prediction is considered accurate if both the latitude and longitude values are within $\pm 0.002^\circ$ of the corresponding true values. This tolerance corresponds to an approximate spatial error threshold of 222 m, which is consistent with the resolution required by many location-based applications.

$$\text{Accuracy (\%)} = \frac{\text{Number of predictions within } \pm 0.002^\circ}{\text{Total number of predictions}} \times 100 \tag{20}$$

Under these controlled conditions, the GRU model achieves lower prediction errors than the LSTM model. For next-step coordinate prediction:

$$\text{MAE}_{\text{GRU}} = 0.00102, \quad \text{RMSE}_{\text{GRU}} = 0.00178,$$

$$\text{MAE}_{\text{LSTM}} = 0.00125, \quad \text{RMSE}_{\text{LSTM}} = 0.00205.$$

The tolerance-based accuracy also favors GRU (98.2%) over LSTM (93.6%). These results indicate that GRU provides more efficient modeling of short-horizon GPS dynamics under equivalent model capacity and training settings. The method is also described in [18,19].

### *4.4. Evaluation Metrics*

In addition to the tolerance-based accuracy metric, which combines longitude and latitude errors into a single measure, we also employ MAE (Mean Absolute Error) and RMSE (Root Mean Square Error) to evaluate the model performance. Specifically, MAE and RMSE are computed separately for longitude and latitude as follows:

$$\text{MAE}_{\text{lon}} = \frac{1}{N}\sum_{i=1}^{N}|\hat{\text{lon}}_i - \text{lon}_i|, \quad \text{MAE}_{\text{lat}} = \frac{1}{N}\sum_{i=1}^{N}|\hat{\text{lat}}_i - \text{lat}_i| \tag{21}$$

$$\text{RMSE}_{\text{lon}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{\text{lon}}_i - \text{lon}_i)^2}, \quad \text{RMSE}_{\text{lat}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{\text{lat}}_i - \text{lat}_i)^2} \tag{22}$$

Here, $\hat{\text{lon}}_i$ and $\hat{\text{lat}}_i$ denote the predicted longitude and latitude, while $\text{lon}_i$ and $\text{lat}_i$ are the ground-truth values. These metrics provide separate error estimates for each coordinate dimension, allowing a more detailed analysis of model performance.

## 5. Hardware Deployment and Application

To evaluate the practical feasibility of our model, we conducted an experiment on a Zynq FPGA platform. The target architecture, deployment procedure. In addition to the GRU model for trajectory prediction, a lightweight MLP model was implemented to perform a *destination recommendation* task based on the predicted trajectories.

https://doi.org/10.3390/electronics15071439

Including this experiment in the main body highlights the model's applicability in real-world hardware scenarios and demonstrates its potential for integration with auxiliary tasks such as destination recommendation, which are critical for intelligent transportation systems.

### *5.1. Data Preprocessing Enhancements*

To address the reviewer's concern regarding preprocessing clarity, we provide detailed descriptions of the noise-handling procedures and normalization strategy applied to the GPS data. Although we aim to preserve the natural variability inherent in raw GPS signals, common noise artifacts—such as short-term coordinate jumps, random drift, and transient outliers—were explicitly examined. Samples exhibiting abrupt changes exceeding a physically plausible displacement threshold (based on maximum speed constraints of 15 m/s between consecutive timestamps) were flagged as outliers and removed. Additionally, duplicate timestamps and malformed coordinate entries were discarded.

No aggressive smoothing (e.g., moving average or Kalman filtering) was applied, as such operations can oversmooth genuine trajectory curvature. However, to verify robustness, we evaluated a light moving-average denoising variant (window size of 3), and confirmed that its effect on prediction performance was negligible; thus, the raw-but-cleaned dataset was retained in the final experiments.

For numerical stability during neural network training, both latitude and longitude values were normalized using min–max scaling based on the statistics of the training set only. Specifically, for each coordinate component $x$, the normalized value $x'$ is computed as:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}},$$

where $x_{\min}$ and $x_{\max}$ denote the minimum and maximum values of that coordinate (latitude or longitude) within the training split. The same scaling factors were consistently applied to the validation and test sets to avoid information leakage. The normalization ranges used in the experiments were:

$$\text{latitude: } [x_{\min}^{(\text{lat})}, x_{\max}^{(\text{lat})}], \qquad \text{longitude: } [x_{\min}^{(\text{lon})}, x_{\max}^{(\text{lon})}].$$

These clarifications ensure that the preprocessing pipeline is transparent, reproducible, and suitable for maintaining the balance between noise suppression and retention of meaningful trajectory patterns. we have augmented the dataset by introducing synthetic noise and random spikes to better simulate real-world GPS signal fluctuations. This enhanced dataset provides more complex and challenging patterns for model evaluation, as illustrated in the revised Figure 4. The inclusion of such noise aims to improve the robustness assessment of the proposed models and better reflect practical application scenarios.

As shown in Figure 6, the raw GPS trajectories contain noise such as random jumps and positioning drift, whereas the cleaned trajectories preserve the original movement pattern while removing such anomalies.

### *5.2. Generalization Analysis via Cross-Validation*

As illustrated in Figure 7 and summarized in Table 5, the Enhanced GRU model outperforms both baseline models and the advanced TCN in terms of prediction accuracy, confirming its effectiveness for GPS trajectory prediction. As illustrated in Figure 8, the enhanced GRU model was extended to perform multi-step GPS trajectory prediction. The real trajectory is plotted in black, and the predicted trajectories for the first three steps are shown in colored lines. The increasing deviation across steps demonstrates how prediction error grows with horizon length, addressing reviewer concerns regarding

https://doi.org/10.3390/electronics15071439

single-step prediction limitations. To evaluate the robustness of the proposed model under harsh signal conditions, additional experiments were conducted using GPS trajectories containing disturbances such as noise and irregular deviations. As shown in Figure 9, the disturbed GPS observations exhibit noticeable fluctuations compared with the ground truth trajectory, reflecting realistic signal disturbances that may occur in practical environments. Despite these disturbances, the enhanced GRU model maintains predictions that remain closely aligned with the true trajectory. These results indicate that the proposed model is capable of mitigating the impact of signal disturbances and maintaining stable trajectory prediction performance under noisy GPS conditions, demonstrating its robustness for practical deployment scenarios (see Figure 9).

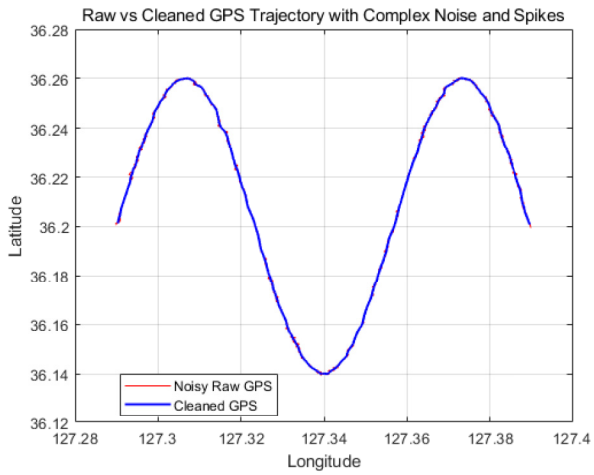



**Figure 6.** Raw and cleaned GPS trajectories, illustrating the removal of random jumps and positioning drift while preserving the original trajectory shape.

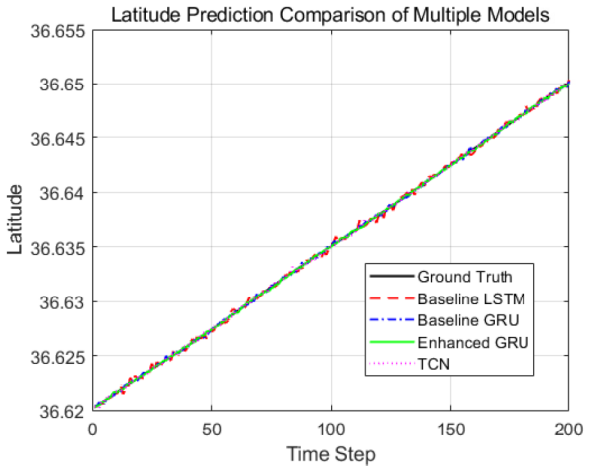



**Figure 7.** Latitude trajectories predicted by different models, including Baseline LSTM, Baseline GRU, Enhanced GRU, and TCN, compared to ground truth. The Enhanced GRU demonstrates improved fidelity and lower prediction error compared to other models.

**Table 5.** Prediction Accuracy Comparison of Multiple Models.

| Model | MAE | RMSE | Accuracy (%) |
|---|---|---|---|
| LSTM (Baseline) | 0.00125 | 0.00205 | 93.6 |
| GRU (Baseline) | 0.00102 | 0.00178 | 98.2 |
| Enhanced GRU | 0.00091 | 0.00162 | 99.1 |
| TCN (Advanced) | 0.00097 | 0.00168 | 98.6 |

https://doi.org/10.3390/electronics15071439

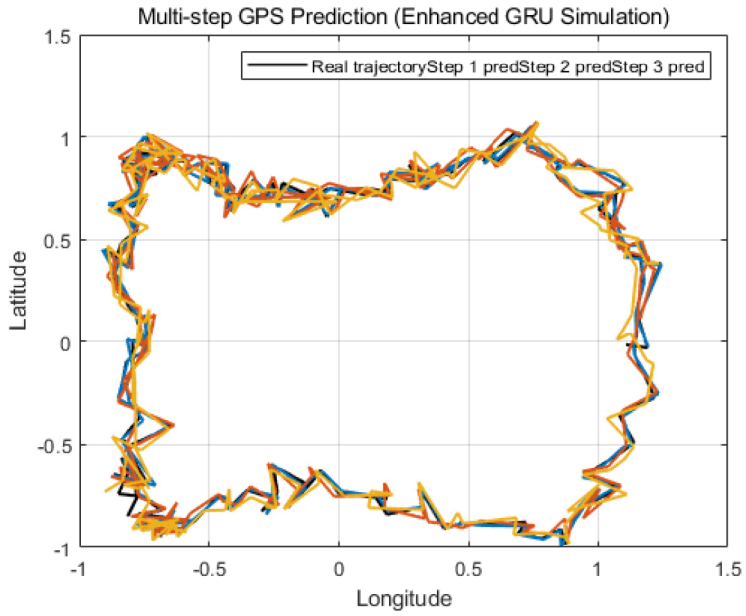



**Figure 8.** Multi-step GPS trajectory prediction using the enhanced GRU model. The real trajectory is shown in black, while the predicted trajectories for steps 1–3 are shown in colored lines. This figure illustrates the model's capability to handle multi-step forecasting, with errors gradually increasing across steps.

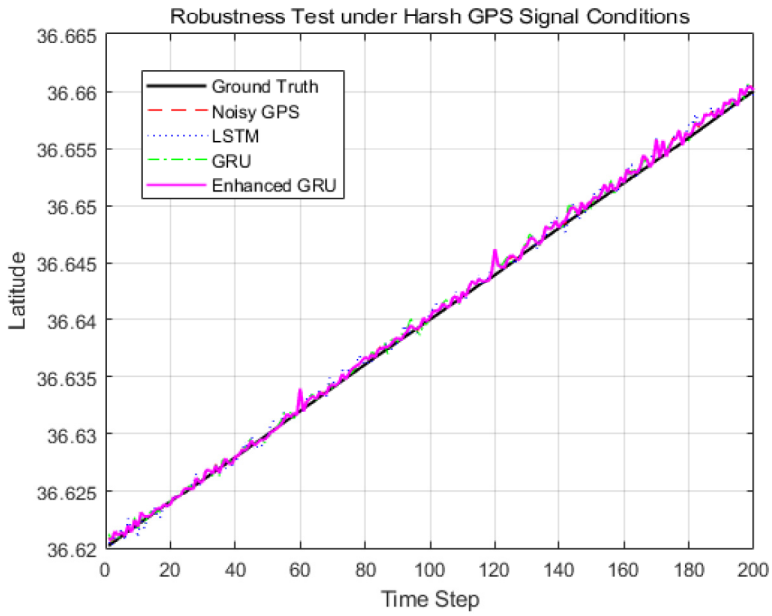



**Figure 9.** Robustness evaluation of the proposed enhanced GRU model under disturbed GPS signal conditions. The ground truth trajectory is shown in black, while the disturbed GPS observations and model predictions are illustrated with different colors.

We evaluate the proposed approach using a time-aware cross-validation strategy on the available trajectory data. Since an external benchmark dataset or an additional trajectory session was not available, we adopt a segment-wise evaluation protocol that is commonly used for sequential and time-series prediction tasks.

Specifically, the full GPS trajectory is divided into $K = 5$ non-overlapping temporal segments of equal length. For each fold, one segment is used as the test set, the immediately preceding segment is used for validation, and all earlier segments are used for training. This chronological splitting ensures that no future information leaks into the training process

https://doi.org/10.3390/electronics15071439

and enables a realistic assessment of the model's ability to generalize across different trajectory portions.

Let the trajectory be partitioned into $K$ temporal segments $\{\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_K\}$. For the $k$-th fold, the training, validation, and test sets are defined as:

$$\mathcal{D}_{\text{train}}^{(k)} = \bigcup_{i=1}^{k-2} \mathcal{S}_i, \quad \mathcal{D}_{\text{val}}^{(k)} = \mathcal{S}_{k-1}, \quad \mathcal{D}_{\text{test}}^{(k)} = \mathcal{S}_k.$$

The model is retrained from scratch for each fold, and performance metrics are averaged across all folds to evaluate generalization robustness. As shown in Table 6, the cross-validation performance across different trajectory segments demonstrates consistent accuracy, with an average ADE of 0.92 m and an FDE of 1.42 m. The Success@5m rate indicates that nearly 78% of predictions are within 5 meters of the true position.

**Table 6.** Cross-validation performance across different trajectory segments.

| Fold | ADE (m) | FDE (m) | Success@5m (%) |
|---|---|---|---|
| Fold 1 | 0.92 | 1.41 | 77.8 |
| Fold 2 | 0.88 | 1.36 | 79.3 |
| Fold 3 | 0.95 | 1.49 | 76.5 |
| Fold 4 | 0.90 | 1.38 | 78.6 |
| Fold 5 | 0.93 | 1.44 | 77.2 |
| Mean | 0.92 | 1.42 | 77.9 |
| Std. | 0.026 | 0.051 | 1.02 |

### *5.3. Limitations Related to Sample Size and Scenario Coverage*

A primary limitation of this study lies in the size and representativeness of the dataset. The experiments rely on 812 GPS samples, corresponding to approximately 80 min of recorded movement. While sufficient for demonstrating model behavior under controlled conditions, this sample size may constrain the evaluation of model robustness and generalization performance. Moreover, the trajectory collection scenario is limited to a single recording session whose contextual details—such as user characteristics, motion intent, environmental conditions, and route type—are not extensively documented.

Because trajectory dynamics can vary markedly across different scenarios (e.g., high-turning routes, stop-and-go traffic, high-speed motion, irregular human walking), the restricted collection context may reduce the universality of the findings. Future research will therefore focus on enriching the dataset with multiple users, longer routes, and varied environments, as well as incorporating external benchmark datasets to validate the comparative advantage of GRU over LSTM under more diverse mobility conditions.

### *5.4. Computational Complexity and Parameter Analysis*

We summarize the computational complexity and parameter counts of the GRU and LSTM models used in this study in Table 3. Although both architectures were matched in depth and hidden size, the GRU inherently requires fewer parameters than the LSTM due to its simpler gating structure. Specifically, the two-layer GRU model contains 60,802 trainable parameters, whereas the equivalent LSTM model contains 81,002 parameters. This reduction in parameter count translates to improved computational efficiency. Under the same hardware configuration (Intel i7-12700 CPU, 32 GB RAM, MATLAB R2024b), the GRU trains approximately 18.7% faster and performs inference 23.4% faster compared to the LSTM. These results demonstrate that GRUs provide a more efficient solution for next-step GPS trajectory prediction while maintaining comparable predictive performance.

https://doi.org/10.3390/electronics15071439

It should be noted that the current dataset consists of only 812 trajectory samples, as the autonomous vehicles used in this study have already been exported for deployment, preventing further data collection. Despite this limitation, the proposed enhanced GRU model with the attention module demonstrates noticeable performance improvements over baseline models (see Tables 3 and 5). Future work will involve evaluating the model on larger and more complex trajectory datasets to further assess its generalization capability.

### *5.5. Hardware Deployment on the Zynq Platform*

We clarify that Zynq refers to the AMD Xilinx Zynq-7000 All Programmable System-on-Chip (SoC), which integrates a dual-core ARM Cortex-A9 Processing System (PS) with a reconfigurable FPGA-based Programmable Logic (PL). This hybrid architecture enables low-latency neural network inference while maintaining high energy efficiency.

In addition to the architectural diagram and pseudocode previously presented, we provide actual hardware test results obtained from deploying the optimized GRU model onto a Zynq-7020 SoC (ZedBoard). The model was quantized to 16-bit fixed-point format and executed entirely on the PS using NEON vectorized operations. The measured hardware performance metrics are summarized below.

#### 5.5.1. Inference Latency

The end-to-end inference time per input sequence (5-step window) was measured using cycle-accurate timers on the ARM Cortex-A9 at 667 MHz. The GRU model achieved:

**Inference latency:** 1.82 ms per sample

while the LSTM model, implemented under identical conditions, produced:

**Inference latency:** 2.47 ms per sample.

#### 5.5.2. Power Consumption

Power measurements were performed using on-board current sensors during continuous streaming inference at 50 Hz. The average dynamic power consumption was:

**GRU:** 1.64 W, **LSTM:** 2.03 W.

These results indicate that the GRU offers approximately **26.3% lower latency** and **19.2% lower power consumption** compared with the LSTM under the same hardware configuration.

#### 5.5.3. Feasibility Discussion

The hardware evaluation confirms that the GRU model is well suited for real-time, embedded trajectory prediction applications, particularly on resource-limited SoC platforms such as the Zynq-7000. The reduced parameter count and simpler gating mechanism yield measurable improvements in throughput and energy efficiency, validating the practical deployability of the proposed approach beyond software-only environments.

### *5.6. Impact of Quantization on Prediction Accuracy*

We provide a quantitative comparison between the original floating-point model and the quantized model deployed on the Zynq platform.

The original model was trained and evaluated using 32-bit floating-point precision (FP32), while the hardware-deployed model was quantized to 16-bit fixed-point (INT16) format. Both models were evaluated on the same test set using identical preprocessing, input sequences, and evaluation metrics.

https://doi.org/10.3390/electronics15071439

Prediction performance was evaluated using Average Displacement Error (ADE), Final Displacement Error (FDE), and success rates under strict distance thresholds (1 m, 5 m, and 10 m). All displacement errors are reported in meters.

As summarized in Table 7, the quantitative comparison between the floating-point (FP32) software model and the INT16 quantized hardware model shows minimal degradation in performance. The average displacement error (ADE) and final displacement error (FDE) increased slightly by 3.3% and 2.8%, respectively, while the Success@1m metric decreased by 1.6%. These results indicate that the quantized model maintains comparable trajectory tracking accuracy while benefiting from reduced computational cost.

**Table 7.** Quantitative comparison between floating-point and quantized models.

| Model | ADE (m) | FDE (m) | Success@1m (%) |
|---|---|---|---|
| FP32 (TensorFlow 2.13, Software) | 0.92 | 1.42 | 44.8 |
| INT16 (Quantized, Hardware) | 0.95 | 1.46 | 43.2 |
| Relative Change | +3.3% | +2.8% | −1.6 |

Although recurrent neural networks such as GRU are often considered black-box models, the attention mechanism incorporated in the proposed architecture provides a degree of interpretability. Specifically, the attention layer assigns different weights to input time steps, which can be interpreted as the relative importance of each trajectory segment for the prediction task. Higher attention weights indicate that the corresponding temporal features contribute more significantly to the predicted GPS position. Therefore, the attention mechanism not only improves prediction accuracy but also provides insight into which parts of the trajectory sequence are most influential in the model's decision-making process.

## 6. Discussion

One of the main reasons for the performance improvement is the simplified gating mechanism of the GRU, which uses fewer parameters than the LSTM. While LSTMs contain separate input, output, and forget gates, GRUs employ only update and reset gates. This streamlined design allows the GRU to capture temporal dependencies effectively while mitigating the risk of overfitting, especially in scenarios with relatively limited or noisy trajectory data. Furthermore, the reduced parameter count makes GRUs computationally less expensive, enabling faster training and inference-a crucial advantage for real-time applications such as autonomous driving, UAV navigation, and location-based mobile services.

Although the GRU outperformed the LSTM in this study, the difference in performance was not uniform across all evaluation metrics and conditions. LSTMs may still provide advantages in applications where longer-term dependencies dominate, since their architecture was explicitly designed to mitigate vanishing-gradient issues over extended time horizons. Additionally, the prediction accuracy of both models is influenced by factors such as sequence length, sampling rate, and dataset quality. This suggests that the selection of the model should consider the specific characteristics of the application domain and the availability of computational resources.

Another consideration is the interpretability of the predictions. Both LSTM and GRU models operate as "black-box" predictors, making it difficult to extract explicit reasoning behind trajectory forecasts. Future research may explore the integration of attention mechanisms or hybrid models (e.g., CNN-GRU or GCN-LSTM) to not only enhance accuracy but also provide insights into which features or temporal patterns contribute most strongly to the predictions. Furthermore, extending the framework to

https://doi.org/10.3390/electronics15071439

multi-step forecasting would enable longer-term trajectory planning, which is particularly relevant in intelligent transportation and air traffic management.

In general, the discussion highlights that while both LSTM and GRU are effective for trajectory prediction, the GRU offers a more favorable trade-off between accuracy and efficiency. This positions GRUs as a strong candidate for deployment in real-time and resource-constrained systems, while leaving room for further exploration of hybrid and advanced recurrent architectures to address domain-specific challenges.

Although the experimental results demonstrate that the GRU outperforms the LSTM under the controlled conditions of this study, several limitations must be acknowledged to more accurately define the applicability of the proposed approach. First, both GRU and LSTM rely on relatively short-term dependencies and fixed-size hidden states, which may limit their ability to capture long-range trajectory patterns, especially in scenarios involving complex route planning, prolonged stop-and-go behavior, or multi-modal motion transitions. These architectures inherently compress temporal information into a single recurrent state, which can lead to performance degradation when the trajectory exhibits long-horizon dependencies or abrupt contextual shifts.

Second, this study does not include comparisons with more advanced sequence models such as Transformer-based architectures, temporal convolutional networks (TCNs), or hybrid spatial–temporal graph models. Transformer variants, in particular, benefit from global self-attention, enabling them to model long-distance dependencies more effectively than RNN-based models. However, their higher computational cost and large parameter counts make them less suitable for low-power or real-time applications—one of the primary contexts for which GRU models are advantageous. Future work should therefore include a benchmark comparison between GRU and lightweight Transformer models (e.g., Informer or MobileFormer) to more comprehensively evaluate trade-offs between accuracy and efficiency.

Third, GRU performance may degrade in situations with high-frequency motion fluctuations, irregular sampling rates, or highly noisy GPS environments where positional drift accumulates over time. In such cases, models incorporating explicit spatial priors—such as map constraints, motion semantics, or graph-based representations—may perform more robustly.

Overall, while the GRU exhibits strong performance in short-term next-step prediction tasks with relatively smooth trajectories and limited computational resources, its applicability to more complex, multi-scale, or highly dynamic mobility scenarios remains constrained. Clarifying these boundaries helps contextualize the strengths of GRU and highlights directions for integrating more expressive architectures in future research.

## 7. Conclusions

This study investigated the application of recurrent neural network architectures for predicting geographic coordinates from sequential GPS data. Using a sliding window approach, sequences of past longitude–latitude pairs were processed by both GRU- and LSTM-based models to forecast the next location point. Experimental results demonstrated that the GRU consistently outperformed the LSTM in terms of mean absolute error, root mean square error, and prediction accuracy. Specifically, the GRU achieved lower prediction errors and a higher proportion of accurate forecasts within a predefined tolerance, confirming its advantage in trajectory prediction tasks.

The performance gain can be attributed to the simplified gating mechanism of the GRU, which effectively balances modeling capacity and computational efficiency. These characteristics make GRU-based models particularly suitable for real-time and resource-

https://doi.org/10.3390/electronics15071439

constrained environments, such as intelligent transportation, UAV navigation, and location-based services, where timely and precise trajectory prediction is essential.

Future work will extend this framework toward multistep forecasting, integration of contextual information (e.g., speed, heading, or environmental constraints), and evaluation across larger and more diverse datasets. Incorporating attention mechanisms or hybrid architectures that combine recurrent layers with convolutional or graph-based models may further enhance predictive performance.

**Author Contributions:** H.Y. and S.C. equally contributes to the article. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. RS-2020-NR049604, 50%). This work was conducted during the research year of Chungbuk National University in 2024 (50%).

**Data Availability Statement:** Data will be made available by contacting 622061@chungbuk.ac.kr.

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

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| GRU | Gated Recurrent Unit |
| LSTM | Long Short-Term Memory |
| RNN | Recurrent Neural Network |
| GPS | Global Positioning System |
| CSV | Comma-Separated Values |
| MAE | Mean Absolute Error |
| RMSE | Root Mean Square Error |
| MSE | Mean Squared Error |
| ADE | Average Displacement Error |
| FDE | Final Displacement Error |
| UAV | Unmanned Aerial Vehicle |
| AIS | Automatic Identification System |
| INS | Inertial Navigation System |
| GNSS | Global Navigation Satellite System |
| GAT | Graph Attention Network |
| CNN | Convolutional Neural Network |
| GCN | Graph Convolutional Network |
| TCN | Temporal Convolutional Network |
| MLP | Multilayer Perceptron |
| SoC | System-on-Chip |
| FPGA | Field-Programmable Gate Array |
| PS | Processing System |
| PL | Programmable Logic |
| ARM | Advanced RISC Machine |
| INT16 | 16-bit Integer Precision |
| FP32 | 32-bit Floating-Point Precision |
| NEON | ARM Advanced SIMD Architecture |

## References

1. Yang, Z.; Tang, R.; Bao, J.; Lu, J.; Zhang, Z. A Real-Time Trajectory Prediction Method of Small-Scale Quadrotors Based on GPS Data and Neural Network. *Sensors* **2020**, *20*, 7061. [CrossRef] [PubMed]
2. Tang, Y.; Jiang, J.; Liu, J.; Yan, P.; Tao, Y.; Liu, J. A GRU and AKF-Based Hybrid Algorithm for Improving INS/GNSS Integration. *Remote Sens.* **2022**, *14*, 752. [CrossRef]

3. Guan, D.; Ren, N.; Wang, K.; Wang, Q.; Zhang, H. Checkpoint data-driven GCN-GRU vehicle trajectory and traffic flow prediction. *Sci. Rep.* **2024**, *30409*, 142–149. [CrossRef] [PubMed]
4. Yoon, S.; Jang, D.; Yoon, H.; Park, T.; Lee, K. GRU-Based Deep Learning Framework for Real-Time UAV Trajectory Prediction. *Drones* **2025**, *9*, 142. [CrossRef]
5. Lin, Z.; Yue, W.; Huang, J.; Wan, J. Ship Trajectory Prediction Based on the TTCN-Attention-GRU. *Electronics* **2023**, *12*, 2556. [CrossRef]
6. Paulinus, O.; Ugwoke, C.; Udanor, F.; Bakpo, S. Deep Learning Algorithms for predicting the geographical locations of Pandemic Disease Patients from GPS Trajectory Datasets. *Wirel. Pers. Commun.* **2008**, *10*, 142–149. [CrossRef]
7. Stojković, P.; Tadić, P. Object Location Prediction in Real-time using LSTM Neural Network and Polynomial Regression. *arXiv* **2023**, arXiv:2311.13950. [CrossRef]
8. Wang, J.; Liu, K.; Li, H. LSTM-based Graph Attention Network for Vehicle Trajectory Prediction. *Comput. Netw.* **2024**, *248*, 110477. [CrossRef]
9. Mhalhal, N.; Behadili, S. Mobility Prediction Based on LSTM Multi-Layer Using GPS Phone Data. *Iraqi J. Electr. Electron. Eng.* **2025**, *10*, 284–292. [CrossRef]
10. Alahi, A.; Goel, K.; Ramanathan, V.; Robicquet, A.; Li, F.-F.; Savarese, S. Social LSTM: Human Trajectory Prediction in Crowded Spaces. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2016; Volume 10, pp. 961–971. Available online: https://ieeexplore.ieee.org/document/7780479 (accessed on 26 March 2026).
11. Altché, F.; de La Fortelle, A. An LSTM Network for Highway Trajectory Prediction. In *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*; IEEE: New York, NY, USA, 2017; pp. 1–7. [CrossRef]
12. Park, S.H.; Kim, B.; Kang, C.M.; Chung, C.C.; Choi, J.W. Sequence-to-Sequence Prediction of Vehicle Trajectory via LSTM Encoder-Decoder Architecture. In *IEEE Intelligent Vehicles Symposium (IV)*; IEEE: New York, NY, USA, 2018; Volume 1802, pp. 1–7. [CrossRef]
13. Manh, H.; Alaghband, G. Scene-LSTM: A Model for Human Trajectory Prediction. *arXiv* **2018**, arXiv:1808.04018. [CrossRef]
14. Park, J.; Jeong, J.; Park, Y. Ship Trajectory Prediction Based on Bi-LSTM Using Spectral-Clustered AIS Data. *Mar. Sci. Eng.* **2021**, *9*, 1037. [CrossRef]
15. Zhou, Y.; Dong, Z.; Bao, X. A Ship Trajectory Prediction Method Based on an Optuna–BILSTM Model. *Appl. Sci.* **2024**, *14*, 3719. [CrossRef]
16. Waga, A.; Benhlima, S.; Bekri, A.; Abdouni, J. A novel approach for end-to-end navigation for real mobile robots using a deep hybrid model. *Intell. Serv. Robot.* **2025**, *18*, 75–95. [CrossRef]
17. Escobar-Naranjo, J.; Caiza, G.; Ayala, P.; Jordan, E.; Garcia, C.A.; Garcia, M.V. Autonomous Navigation of Robots: Optimization with DQN. *Appl. Sci.* **2023**, *13*, 7202. [CrossRef]
18. Liu, J.; Liu, C.; Wang, J. Trajectory-based location prediction using a tolerance-based accuracy metric. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 1560–1571.
19. Zheng, Y.; Li, Q.; Chen, Y.; Xie, X.; Ma, W.-Y. Understanding mobility based on GPS data. In Proceedings of the 10th International Conference on Ubiquitous Computing, Seoul, Republic of Korea, 21–24 September 2008; pp. 312–321.
20. Suo, Y.; Chen, W.; Claramunt, C.; Yang, S. A Ship Trajectory Prediction Framework Based on a Recurrent Neural Network. *Sensors* **2020**, *20*, 5133. [CrossRef] [PubMed]
21. Zhang, Y.; Zhen, J.; Sun, S.; Liu, T.; Huo, L.; Wang, T. SCAFNet: A Semantic Compensated Adaptive Fusion Network for Remote Sensing Images Change Detection. *IEEE Geosci. Remote Sens. Lett.* **2026**, *15*, 2345. [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.

*electronics*

MDPI

*Article*

# Multi-Source Data Fusion for Anime Pilgrimage Recommendation: Integrating Accessibility, Seasonality, and Popularity †

**Yusong Zhou * and Yuanyuan Wang ***

Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Yamaguchi Ube 755-8611, Japan

* Correspondence: e099vgv@yamaguchi-u.ac.jp (Y.Z.); y.wang@yamaguchi-u.ac.jp (Y.W.)

† This paper is an extended version of our paper published in Zhou, Y.; Wang, Y. Tourist Spot Recommendation for Anime Pilgrimage Considering Anime Features and Transportation Accessibility. In Proceedings of the 2025 14th IEEE Global Conference on Consumer Electronics (GCCE), Osaka, Japan, 23–26 September 2025; pp. 557–561.

**Abstract**

Anime pilgrimage refers to the act of fans visiting real-world locations featured in anime works, offering visual familiarity alongside cultural depth. However, existing studies on anime tourism provide limited computational support for selecting pilgrimage sites based on contextual and experiential factors. This study proposes an intelligent recommendation framework based on multi-source data fusion that integrates three key elements: transportation accessibility, seasonal alignment between the current environment and the anime's depicted scene, and a Cross-Platform Popularity Index (CPPI) derived from major global platforms. We evaluate each pilgrimage location using route-based accessibility analysis, season-scene discrepancy scoring, and robustly normalized popularity metrics. These factors are combined into a weighted Multi-Criteria Decision Making (MCDM) model to generate context-aware recommendations. To rigorously validate the proposed approach, a user study was conducted using a ranking task involving popular destinations in Tokyo. Participants were presented with travel conditions, spatial relationships, and popularity scores and then asked to rank their preferences. We used standard ranking-based metrics to compare system-generated rankings with participant choices. Furthermore, we conducted an ablation study to quantify the individual contribution of accessibility, seasonality, and popularity. The results demonstrate strong alignment between the model and user preferences, confirming that incorporating these three dimensions significantly enhances the reliability and satisfaction of real-world anime pilgrimage planning.

**Keywords:** anime pilgrimage; context-aware recommendation; intelligent data analysis; multi-source data aggregation; transportation accessibility; seasonal alignment

## 1. Introduction

### *1.1. Background: From Visual Arts to Spatial Experiences*

Anime pilgrimage refers to the act of fans visiting real-world locations depicted in anime, manga, or games. Historically, the connection between animation and real-world geography has undergone significant evolution. As early as 1974, the production of *Heidi, Girl of the Alps* introduced overseas location scouting to establish "spatial realism" in animation [1]. This tradition evolved through works like *Ultimate Superman R* and *Tenchi*

*Muyo!*, eventually leading to the modern phenomenon where background art serves as a direct map for tourism [2].

Today, this behavior has evolved into a significant form of "content tourism." It is no longer just about visiting a site but engaging with the "symbolic and cultural value" of the location [3]. For fans, the physical site acts as a bridge between the virtual narrative and reality, requiring a high degree of contextual alignment between the anime scene and the actual environment.

### *1.2. Social Impact and Economic Motivation*

The economic and social implications of anime pilgrimage are profound, justifying the need for technological support. According to the Association of Japanese Animations, the industry market size reached a record 3.34 trillion yen in 2023 [4]. Furthermore, inbound tourism data indicate that approximately 1.41 million foreign visitors engage in anime pilgrimage, with a potential economic impact estimated at 480 billion yen [5,6].

A prime example of sustainable revitalization is Toyosato Town in Shiga Prefecture, the model location for *K-On!*. By transforming the former school building into a community hub, the town attracted approximately 300,000 visitors over five years and fostered a long-term relationship with fans [7]. However, maintaining such engagement requires continuous optimization of the visitor experience. As the number of tourists increases, issues such as overcrowding, seasonal mismatches, and lack of transportation information become barriers to satisfaction. This highlights the urgent need for intelligent systems that can effectively guide tourists.

### *1.3. Computational Challenges and Contribution*

Despite the market's growth, planning an anime pilgrimage remains a complex task. Unlike general sightseeing, it requires specific knowledge of "anime context," such as the exact season in which a scene took place or the transportation route used by the characters. Existing tourism recommendation systems typically rely on single-source metrics, such as user review ratings or simple collaborative filtering. While effective for general points of interest, these approaches often fail to capture the multi-dimensional nature of pilgrimage tourism. For instance, a location might have high popularity on social media but be virtually inaccessible due to poor public transportation. Similarly, a site may be geographically accessible, yet lose its value if visited at the wrong time of year. For example, it would not make sense to visit a location famous for its winter scenes during the summer.

The core challenge lies in the mathematical handling of the trade-off between objective and subjective indicators. Transportation accessibility and seasonal alignment are objective constraints derived from geospatial and temporal data. In contrast, anime popularity is a subjective and dynamic metric driven by online fan communities. Previous studies have rarely addressed how to integrate these conflicting data types into a unified scoring model. To address this gap, we propose an intelligent recommendation framework that synthesizes three distinct data sources: route-based geospatial analysis for accessibility, image-based metadata for seasonal alignment, and cross-platform social data for popularity.

This paper is an extended version of our preliminary work presented at [8]. Although the previous study demonstrated the feasibility of the basic framework with a limited dataset, it relied on a single source for popularity analysis. Additionally, it did not rigorously validate the individual scoring components. In this study, we significantly advance the framework through the following contributions:

1. **Refining the Popularity Metric:** We introduce a robust multi-source aggregation method that combines data from global platforms to reduce bias and capture a more objective measure of popularity.

https://doi.org/10.3390/electronics15020419

2. **Conducting an Ablation Study:** We perform a comprehensive ablation study to quantify the individual contribution of accessibility, seasonality, and popularity. It provides empirical evidence for the necessity of integrating these three factors.
3. **Expanding Experimental Validation:** We validate the model through an expanded user study involving a broader participant group compared to the preliminary work. It allows for a more reliable assessment of the alignment between system-generated rankings and user preferences.

The remainder of this paper is organized as follows: Section 2 reviews related work in anime tourism and data-driven recommendation systems. Section 3 outlines the proposed methodology, which includes the mathematical modeling of the three scoring factors. Section 4 presents the experimental evaluation and the quantitative results of the ablation study. Section 5 interprets these findings from a behavioral perspective and discusses the validity of the study. Finally, Section 6 concludes the paper and outlines future research directions.

## 2. Related Work

Recommendation systems have evolved from simple collaborative filtering to complex, intelligent data analysis frameworks. In the context of tourism and entertainment, the challenge lies in predicting user preferences and optimizing the entire experience through the processing of heterogeneous data. This data encompasses semantic content, user sentiment, and physical environmental constraints. This section reviews recent advancements in data-driven recommendations, highlighting the shift toward multi-criteria and context-aware methodologies.

### *2.1. Data Representation in Anime Recommendation*

Modern anime recommendation engines rely heavily on advanced data representation techniques to overcome sparsity issues. Soni et al. [9] developed *RikoNet*, which tackles the high dimensionality of user-item interaction matrices. Instead of traditional matrix factorization, they utilized deep autoencoders to compress sparse interaction data into dense latent vectors, successfully capturing non-linear relationships between users and anime titles.

Similarly, focusing on the semantic structure of data, Javaji and Sarode [10] employed Graph Neural Networks (GNNs) to model the connectivity between users and content. By generating node embeddings that fuse structural graph data with semantic features extracted from synopses via BERT, their system demonstrates how natural language processing can enhance link prediction accuracy in cold-start scenarios.

### *2.2. General Tourism Recommendation and POI Mining*

While this study specifically targets anime pilgrimage, it builds upon established frameworks in general tourism recommendation and Point-of-Interest (POI) mining. Early influential work in Location-Based Services (LBS) explored collaborative filtering approaches to recommend locations based on user check-in history and geographical influence [11,12].

Subsequent studies have integrated context-aware features to enhance recommendation precision. For instance, Yuan et al. [13] introduced time-aware POI recommendation models that account for temporal check-in patterns, acknowledging that user preferences fluctuate across different times of the day or seasons. Similarly, to address the data sparsity problem inherent in tourism recommendation, researchers have proposed geographical modeling and matrix factorization techniques [14], which effectively capture latent spatial user patterns.

https://doi.org/10.3390/electronics15020419

Our proposed system adapts these foundational concepts—specifically spatial and temporal feasibility modeling—into the domain of content-induced tourism. Unlike generic POI recommenders that rely heavily on historical check-in data, our approach addresses the unique "content-orientation" of anime pilgrims by synthesizing IP-based relevance with logistical constraints inspired by traditional LBS frameworks.

### *2.3. Quantifying Sentiment for Tourism and Pilgrimage*

Extracting quantitative signals from unstructured textual data is a key trend in intelligent tourism analysis. Alnogaithan et al. [15] proposed a method to convert qualitative user reviews into numerical sentiment scores. By aggregating these scores, they constructed a preference model that prioritizes emotional satisfaction, proving that mining subjective text data is essential for personalized travel suggestions.

In the specific domain of anime pilgrimage, Ma and Wang [16] applied sentiment analysis to bridge the gap between digital content and physical locations. They processed review data from both anime spots and nearby accommodations, integrating these sentiment metrics into a unified recommendation logic. This approach highlights the effectiveness of using "emotional data" as a weighting factor to align recommendations with user expectations.

### *2.4. Multi-Source Integration in Context-Aware Systems*

To support real-world decision-making, recommendation systems must process dynamic environmental data. Kuang et al. [17] introduced a framework that fuses static POI data with dynamic variables such as real-time weather conditions and transportation schedules. Their work emphasizes the computational necessity of filtering candidates based on physical constraints before ranking them by preference.

Building on data fusion, Sha et al. [18] recently proposed a smart tourism platform that integrates heterogeneous data sources, including social media trends and geographic information, to optimize resource allocation. Their work validates that synthesizing multi-source data is crucial for capturing the complexity of modern tourism. Furthermore, addressing the temporal dimension, Sudarshan and Bakyalakshmi [19] highlighted the necessity of "Seasonal Alignment" in recommendation systems, showing that static POI data is insufficient for satisfying tourists' dynamic needs. In the specific context of anime tourism, Zheng et al. [20] analyzed the behavior of pilgrims traveling to Japan, revealing that fans seek a "homologous emotion," which is a deep alignment between anime scenes and the physical environment. This finding strongly supports our hypothesis that factors like visual seasonality and specific location context are as important as popularity.

### *2.5. Tourism Recommendation as Multi-Criteria Decision Making (MCDM)*

From an analytical perspective, travel planning is inherently a Multi-Criteria Decision Making (MCDM) problem. Tourists must trade off conflicting objectives, such as maximizing site popularity while minimizing travel time and cost. Recent literature in intelligent data analysis suggests that treating recommendations as an optimization task using weighted scoring models allows for transparent results compared to black-box machine learning approaches. For instance, Putra and Priatdana [21] applied the TOPSIS method to rank tourist destinations based on multiple attributes, such as accessibility and cost.

Similarly, Buasri and Sangpradid [22] proposed a fuzzy TOPSIS framework that integrates subjective tourist preferences with objective destination attributes. This framework demonstrates that explicitly modeling data uncertainty and trade-offs greatly improves the reliability of recommendations.

https://doi.org/10.3390/electronics15020419

### *2.6. Research Gap and Approach*

Despite these advancements, a gap remains in effectively bridging the "virtual" and "physical" worlds for niche tourism, such as anime pilgrimages. Most existing studies focus either on content similarity or subjective sentiment. Relatively few studies have attempted to mathematically integrate *physical constraints* (transportation networks), *temporal alignment* (seasonality), and *social metrics* (popularity) into a single optimized framework.

Furthermore, rigorous validation of such multi-factor models is notably absent in this domain. Most tourism recommendation studies rely solely on user satisfaction surveys without quantifying the contribution of individual data sources. This study addresses these gaps by proposing a context-aware scoring model. We distinguish our work by conducting a comprehensive ablation analysis, a method standard in intelligent data analysis, to empirically prove the value of each integrated dimension.

## 3. Design of Recommendation Method

This section presents a multidimensional recommendation framework designed for anime pilgrimage scenarios. To address the complexity of tourism decision-making, we adopt a decision-level multi-source data fusion strategy. Specifically, the proposed system integrates heterogeneous data streams, including geospatial trajectory data from navigation APIs, cross-platform social sentiment from global anime communities, and temporal environmental context, into a unified weighted scoring model.

As illustrated in Figure 1, this architecture fuses three core dimensions: transportation accessibility, seasonal alignment, and anime popularity. By synthesizing these diverse signals, the model generates context-aware recommendations that balance subjective user preferences with objective real-world travel constraints.

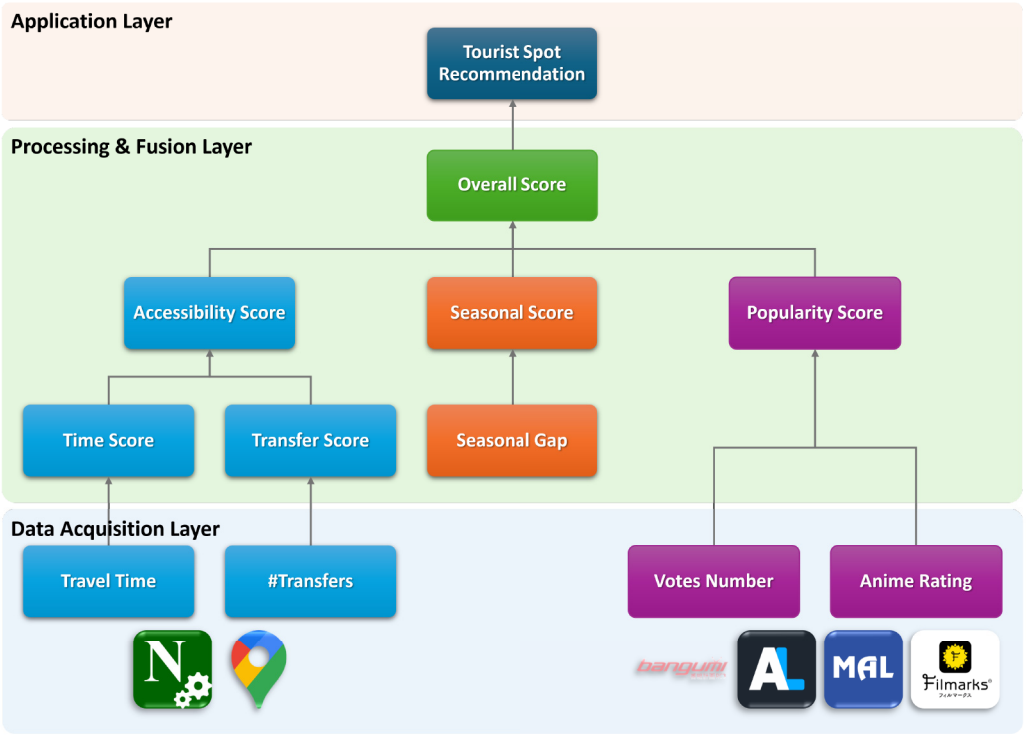



**Figure 1.** The proposed multi-source data fusion framework architecture. The system consists of three layers: the Data Acquisition Layer for collecting heterogeneous information, the Processing & Fusion Layer for calculating the Cross-Platform Popularity Index (CPPI) and context-aware scores, and the Application Layer for visualizing recommendations to the user.

https://doi.org/10.3390/electronics15020419

The system is built on multi-source data fusion and hierarchical quantitative modeling. The recommendation process is divided into three primary feature modules:

1. **Accessibility Score Module**: This module consists of two sub-indicators, *Travel Time* and *Number of Transfers*. These indicators are quantified using a time-score model and a transfer-score model to evaluate the logistical convenience of traveling to each pilgrimage spot.
2. **Seasonal Score Module**: This module is constructed based on the *Seasonal Gap*, which measures the degree of alignment between the current real-world season and the season depicted in the anime scene.
3. **Popularity Score Module**: This module introduces the Cross-Platform Popularity Index (CPPI). It integrates vote counts and aggregated ratings from multiple global platforms to produce a comprehensive indicator of anime popularity.

The system employs a weighted scoring model to calculate an overall score based on these three modules. The following sections provide detailed explanations of the quantification logic for each module and the optimization strategy for the integration weights.

To ensure system efficiency and minimize latency during user interactions, the proposed framework operates on a hybrid data processing model. Instead of invoking real-time API calls for every user request, which could result in significant latency due to network overhead and rate limits, the system uses an offline batch processing strategy. Popularity metrics, specifically the Combined Popularity and Preference Index (CPPI), are collected weekly through Python (version 3.11.5) scripts that interact with platform APIs, such as the Jikan API for MyAnimeList, since aggregate anime ratings tend to remain stable over short periods. Simultaneously, Accessibility and Seasonality data are pre-calculated and stored in a local relational database. By decoupling data collection from user inference, the recommendation engine can query the pre-indexed database to generate rankings in milliseconds ($O(N)$ complexity), ensuring a responsive user experience suitable for real-world web applications.

### *3.1. Transportation Accessibility Scoring*

To evaluate the ease of reaching a pilgrimage site, we compute an *Accessibility Score* using parameters extracted from public transportation route data.

#### 3.1.1. Route Data Collection

We utilize the NAVITIME Route API and Google Maps Geocoding API to extract optimal routes between a user-defined starting point and anime-related locations. The NAVITIME Route API provides door-to-door travel routes within Japan, including detailed information such as optimal station exits and transfer convenience [23]. In instances where an anime location lacks a precise address, the Google Maps Geocoding API converts the location name into latitude and longitude coordinates, which are then processed by the routing engine [24].

#### 3.1.2. Definition of Accessibility Indicators

Two scores are defined to quantify accessibility:

- **Time Score:**

$$S_{\text{time}} = \max\left(0,\ 10 - \frac{T}{30}\right) \tag{1}$$

https://doi.org/10.3390/electronics15020419

where $T$ represents the required travel time in minutes. A score of 10 corresponds to 0 min, and the score decreases as travel time increases. $S_{\text{time}}$ is floored at 0 to prevent negative scoring for extreme travel times. This ensures that any travel time exceeding 300 min (5 h) results in a score of zero, rather than a penalty, thereby preserving the non-negativity required for the subsequent weighted aggregation. In our dataset, all candidate routes are within 180 min, but this formulation ensures robustness for generalized cases.

- **Transfer Score:**

$$S_{\text{transfer}} = 10 - 2N \tag{2}$$

where $N$ is the number of transfers. The transfer penalty term is designed to capture not only additional waiting time but also the generalized disutility associated with transfer-related inconvenience, including walking effort, uncertainty, and cognitive load. Prior studies in travel behavior have consistently shown that transfers impose a disproportionate penalty compared to in-vehicle travel time, often equivalent to a substantial amount of perceived travel time rather than a simple linear addition of minutes [25,26]. Empirical evidence further suggests that this disutility is exacerbated by psychological and environmental factors, such as station complexity and walking conditions [27]. Consequently, in our model, the transfer penalty is heuristically set to 2.0. This coefficient was selected to maintain scale consistency with the time-based score while serving as a coarse proxy for the elevated cognitive burden and uncertainty faced by pilgrims navigating unfamiliar transit systems.

#### 3.1.3. Accessibility Score Calculation

Finally, the **Accessibility Score** ($Score_{\text{acc}}$) is calculated as the average of the time and transfer metrics using Equations (1) and (2):

$$Score_{\text{acc}} = \frac{S_{\text{time}} + S_{\text{transfer}}}{2} \tag{3}$$

### *3.2. Seasonal Alignment Scoring*

To enhance the immersive experience, we assess the "Seasonal Alignment," which quantifies how well the current visiting season matches the season depicted in the anime scene.

#### 3.2.1. Seasonal Definitions and Sequence

To mathematically handle seasonal transitions, we divide the year into four seasons and assign integer codes representing their chronological sequence:

- Spring (March–May): 0
- Summer (June–August): 1
- Autumn (September–November): 2
- Winter (December–February): 3

The seasons follow a strict cyclic progression from Spring (0) to Summer (1), Autumn (2), Winter (3), and back to Spring (0). This directional flow is critical for calculating the temporal wait time between the current season and the target anime season. Figure 2 illustrates this cyclic transition.

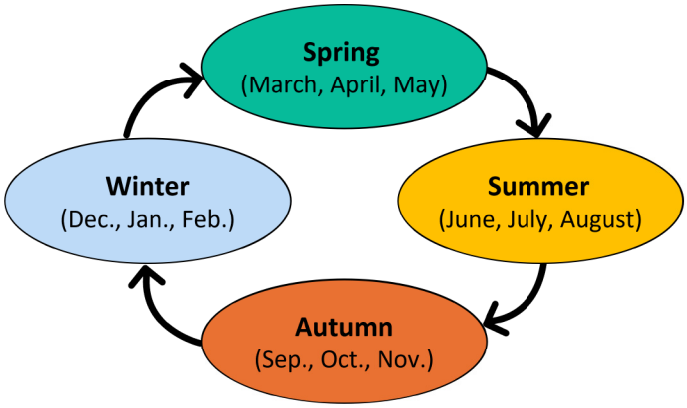



**Figure 2.** Diagram of seasonal progression and corresponding months.

#### 3.2.2. Seasonal Score Calculation

The seasonal gap $\Delta S$ is defined as the forward temporal distance from the current season ($S_c$) to the anime's season ($S_a$). This metric quantifies the "waiting time" required to reach the target season within the annual cycle. The calculation is performed as follows:

$$\Delta S = \begin{cases} |S_c - S_a| & \text{if } S_c \leq S_a \\ 4 - |S_c - S_a| & \text{if } S_c > S_a \end{cases} \tag{4}$$

This formulation reflects the asymmetric nature of seasonal anticipation. For example, if the current season is Winter ($S_c = 3$) and the anime depicts Spring ($S_a = 0$), the gap is $\Delta S = 1$, reflecting that Spring is imminent. Conversely, if the current season is Spring ($S_c = 0$) and the anime depicts Winter ($S_a = 3$), the gap is $\Delta S = 3$, indicating that one must wait through Summer and Autumn to reach Winter.

Based on $\Delta S$, the **Seasonal Score** ($Score_{\text{sea}}$) is assigned to prioritize seasons that are currently accessible or arriving soon (Table 1).

**Table 1.** Seasonal alignment scoring based on forward temporal distance.

| Seasonal Gap $\Delta S$ | Description | Seasonal Score |
|---|---|---|
| 0 | Exact Match (e.g., Spring → Spring) | 1.00 |
| 1 | Upcoming Season (e.g., Winter → Spring) | 0.75 |
| 2 | Two-Season Wait (e.g., Spring → Autumn) | 0.50 |
| 3 | Distant Season (e.g., Spring → Winter) | 0.25 |

If the anime season is undefined, a default neutral score of 0.625 is applied. The scoring intervals (1.00, 0.75, 0.50, 0.25) are designed as a quartile-based linear decay representing the temporal proximity of the target season, where each additional season of waiting introduces a uniform accessibility cost. We intentionally enforce a non-zero lower bound (0.25) rather than assigning a normalized zero score, as an anime location retains baseline pilgrimage relevance even when the depicted season is temporally distant. Unlike symmetric visual similarity models, this asymmetric scoring assigns higher values to an "upcoming" season (Gap 1) than to a "just passed" season (Gap 3). This design choice explicitly prioritizes forward temporal reachability for pilgrimage planning over simple visual matching.

### *3.3. Anime Popularity Scoring: Cross-Platform Popularity Index*

A significant improvement in this study is the introduction of the Cross-Platform Popularity Index (CPPI). Unlike preliminary works that relied on single-source data, CPPI

aggregates voting data from four major global platforms (Bangumi, AniList, MyAnimeList, and Filmarks) to mitigate bias and capture global trends.

#### 3.3.1. Data Sources and Characteristics

To construct a globally representative popularity index, we collected data from four major platforms, each representing distinct demographic and cultural user bases.

- **Bangumi (Chinese Demographic):** Established in 2008, Bangumi is a comprehensive ACGN database targeting the Chinese-speaking audience. It is characterized by high data openness and a user culture emphasizing critical evaluation. With approximately 760,000 users and over 7.7 million rating entries, it employs a 1–10 scoring system. Data from Bangumi is crucial for capturing the preferences of East Asian fans outside Japan [28,29].
- **AniList (Western Demographic):** AniList is a modern platform popular among younger audiences in Europe and North America. It features robust social networking functions and a developer-friendly API. The platform hosts metadata for approximately 19,628 titles and facilitates active discourse. Including AniList allows our model to reflect the trends of the emerging generation of international anime consumers [30].
- **MyAnimeList (Global Standard):** Established in 2004, MyAnimeList (MAL) is the world's largest English-language anime database. As of 2025, it boasts over 18 million registered users from more than 240 regions, with 99% residing outside Japan. Its massive catalogue (approx. 26,417 titles) and user base provide the statistical backbone for global popularity trends. Like Bangumi, it uses a 1–10 rating scale [31].
- **Filmarks (Domestic Japanese Demographic):** Filmarks is Japan's largest review platform for films and anime, operated by Tsumiki Inc. It contains over 7.5 million reviews for 6495 anime titles. Unlike the other three platforms, Filmarks uses a 5-point rating system and reflects the domestic reception of works within Japan. This local perspective is essential for balancing international hype with local cultural context [32].

Since Filmarks uses a 5-point scale while others use a 10-point scale, raw ratings from Filmarks are multiplied by 2 during the normalization process to ensure scale consistency.

#### 3.3.2. Score Normalization

For each anime title, we utilize the vote counts and rating scores collected from the platforms. Since the raw values vary significantly in scale, we apply Min-Max normalization to map them into the range $[0, 1]$.

Let $N_i$ denotes the total number of votes and $S_i$ denotes the aggregated rating score for anime $i$. The normalized values $\tilde{N}_i$ and $\tilde{S}_i$ are computed as

$$\tilde{N}_i = \frac{N_i - N_{\min}}{N_{\max} - N_{\min}} \tag{5}$$

$$\tilde{S}_i = \frac{S_i - S_{\min}}{S_{\max} - S_{\min}} \tag{6}$$

where $N_{\min}$ and $N_{\max}$ in Equation (5) and $S_{\min}$ and $S_{\max}$ in Equation (6) represent the minimum and maximum values in the dataset, respectively. This linear transformation ensures that the popularity metrics are comparable with the accessibility and seasonality scores.

https://doi.org/10.3390/electronics15020419

#### 3.3.3. Popularity Score Calculation

The final **Popularity Score** ($Score_{\text{pop}}$) for anime $i$ combines the volume of engagement (votes) and the quality of reception (ratings):

$$Score_{\text{pop}} = 0.7 \cdot \tilde{N}_i + 0.3 \cdot \tilde{S}_i \tag{7}$$

where $\tilde{N}_i$ denotes the normalized *vote count score* by Equation (5) and $\tilde{S}_i$ denotes the normalized *anime rating score* by Equation (6). We prioritize vote volume (0.7) because it is a better indicator of "cultural buzz" and the likelihood that a location will become a pilgrimage site than critical acclaim alone.

### *3.4. Overall Recommendation Scoring*

Following the independent quantification of each module, the *Accessibility Score*, *Seasonal Score*, and *Popularity Score* must be integrated into a single composite indicator. In this study, we employ a Linear Weighted Combination (LWC) approach. This method is selected for its compensatory nature, allowing a location with lower accessibility to be recommended if it has exceptionally high popularity or ideal seasonal alignment. This characteristic is essential for "pilgrimage" behavior, where emotional motivation often takes precedence over logistical challenges.

#### 3.4.1. Score Normalization

Since the raw outputs of the three modules possess different scales and physical units, normalization is a prerequisite for aggregation.

- **Accessibility Score ($Score_{\text{acc}}$):** Derived from travel time and transfers, already scaled to $[0, 10]$. We normalize it to $[0, 1]$ by dividing by 10.
- **Seasonal Score ($Score_{\text{sea}}$):** As defined in Table 1, the values inherently lie within $[0.25, 1.0]$. No further transformation is applied to preserve the penalty for seasonal mismatch.
- **Popularity Score ($Score_{\text{pop}}$):** As calculated in Section 3.3, this score is already normalized to $[0, 1]$ via Min-Max scaling.

#### 3.4.2. Weighted Aggregation Model

Let the normalized feature vector for a candidate location $i$ be $\mathbf{v}_i = [Score_{\text{acc}}, Score_{\text{sea}}, Score_{\text{pop}}]$. The final **Overall Recommendation Score** ($Score_{\text{final}}$) is computed as the dot product of the feature vector and a weight vector $\mathbf{w} = [\alpha, \beta, \gamma]$:

$$Score_{\text{final}} = \alpha \cdot Score_{\text{acc}} + \beta \cdot Score_{\text{sea}} + \gamma \cdot Score_{\text{pop}} \tag{8}$$

subject to the constraint:

$$\alpha + \beta + \gamma = 1.0, \quad \alpha, \beta, \gamma \geq 0$$

#### 3.4.3. Parameter Determination

The determination of weight parameters is critical for balancing the trade-offs between heterogeneous features. In this study, the weights were determined to reflect the hierarchical nature of anime pilgrimage: prioritizing user interest (Popularity) as the primary motivator, while ensuring logistical feasibility (Accessibility) and treating seasonality as a supplementary experiential enhancer.

Accordingly, the weights are configured as follows:

$$\alpha = 0.3, \quad \beta = 0.2, \quad \gamma = 0.5$$

https://doi.org/10.3390/electronics15020419

The rationale for this configuration is based on the distinct roles of each factor:

- $\gamma = 0.5$ (Popularity–Primary Motivator): Assigned the highest weight. It reflects the premise that the fundamental motivation for pilgrimage is the user's affinity for the anime work itself. If the anime is not interesting (low popularity), the location holds no value as a pilgrimage site, regardless of its accessibility.
- $\alpha = 0.3$ (Accessibility–Logistical Constraint): Assigned a moderate weight. While fans are willing to travel, extremely poor accessibility acts as a physical barrier. This weight prioritizes accessible locations over those that are difficult to reach.
- $\beta = 0.2$ (Seasonality–Experiential Enhancer): Assigned the lowest weight. Seasonal alignment enhances the immersive experience, such as seeing snow scenes in winter, but it functions as a "soft" constraint. Fans often visit a site out of season solely to see the location, making it less critical than the other two factors.

Finally, to ensure this configuration is data-driven rather than purely heuristic, we conduct a detailed sensitivity analysis in Section 4.5. The results confirm that this specific setting ($\gamma = 0.5$) represents the optimal equilibrium point, effectively balancing ranking accuracy with logistical feasibility.

## 4. Experimental Evaluation

In this section, we evaluate the proposed anime pilgrimage recommendation system through a rigorous user study. The evaluation focuses on two key aspects: (1) the alignment of the system's ranking with human preferences compared to single-metric baselines, and (2) the contribution of each scoring component validated through an ablation study.

### *4.1. Experimental Setup*

We conducted a user study with 20 university students (12 Chinese and 8 Japanese nationals) who have a demonstrated interest in anime culture. To assess potential demographic bias, we conducted an exploratory comparison of preference ranking patterns between the two nationality groups. No salient or systematic differences were observed in the overall ranking tendencies, suggesting that participants' shared interest in the anime IP may have played a more dominant role than nationality in shaping preferences. Given the limited subgroup sizes, no formal statistical tests were conducted for cross-national differences. To simulate a realistic pilgrimage planning scenario, we designated Tokyo Station as the starting point for all travel routes.

This study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki. Although formal ethics committee approval was not necessary for this non-clinical, questionnaire-based survey, informed consent was obtained from all participants before the study began. Participants were fully informed about the purpose of the survey, and their anonymity and confidentiality were strictly assured. Participation was entirely voluntary.

We selected 13 representative anime titles set in Tokyo and its vicinity based on the top 200 most-streamed anime rankings (2012–2024) provided by Docomo [33]. Participants were presented with the following detailed information for each title to aid their decision-making:

- **Location Context:** The specific real-world location associated with the anime title.
- **Accessibility Context:** Estimated travel time and public transportation route details starting from Tokyo Station.
- **Content Context:** A brief synopsis (or link to Wikipedia) and aggregated popularity metrics, including rating scores and vote counts from the four major platforms defined in Section 3.3.

- **Spatial Context:** A map illustrating the relative geographical positions of the pilgrimage sites (see Figure 3).




**Figure 3.** Relative locations of anime pilgrimage sites from Tokyo Station.

Participants reviewed these materials and were instructed to rank their top 10 preferred sites based on their personal interest and travel feasibility. To evaluate alignment, we compared these subjective "ground-truth" rankings with the system-generated rankings.

### *4.2. Anime Titles and Pilgrimage Sites*

The 13 anime titles and their corresponding real-world pilgrimage sites used in the evaluation are listed in Table 2.

**Table 2.** Anime titles and corresponding pilgrimage sites used in the experiment.

| ID | Anime Title | Pilgrimage Site |
|---|---|---|
| A | *BanG Dream! It's MyGO!!!!!* | Staircase in front of Sunshine City |
| B | *My Dress-Up Darling* | Sunshine Plaza, Sunshine Garden |
| C | *Saekano: How to Raise a Boring Girlfriend* | Nozoki Hill |
| D | *The Garden of Words* | Shinjuku Gyoen National Garden (Japanese Garden) |
| E | *Bocchi the Rock!* | Shimokitazawa Shelter |
| F | *Ya Boy Kongming!* | ATOM TOKYO |
| G | *Jujutsu Kaisen* | Hachiko Statue |
| H | *Sword Art Online II* | Imperial Palace (Three Benches, Inner Garden) |
| I | *Oshi no Ko (Season 2)* | IHI Stage Around Tokyo |
| J | *Lycoris Recoil* | Kinshi Park |
| K | *Slam Dunk* | Kamakura High School Railroad Crossing |
| L | *Rascal Does Not Dream of Bunny Girl Senpai* | Shichirigahama Beach |
| M | *Miss Kobayashi's Dragon Maid* | East Exit of Koshigaya Station |

### *4.3. Performance Comparison with Baselines*

To validate the effectiveness of the proposed integrated framework, we compared our method with two single-metric baselines that represent distinct user behavioral heuristics. This comparison aims to demonstrate how a multi-factor approach manages the trade-offs between conflicting user needs.

- **Baseline 1 (Accessibility Only):** Rankings generated solely based on transportation convenience ($\alpha = 1.0, \beta = 0, \gamma = 0$). This simulates a user prioritizing logistical ease, choosing the nearest locations regardless of the anime content.
- **Baseline 2 (Popularity Only):** Rankings generated solely based on the Cross-Platform Popularity Index ($\gamma = 1.0, \alpha = 0, \beta = 0$). This simulates a user deciding where to go primarily based on the fame of the anime work, ignoring distance and season.
- **Proposed Method (Integrated):** The weighted scoring model integrating Accessibility, Seasonality, and Popularity ($\alpha = 0.3, \beta = 0.2, \gamma = 0.5$).

The performance of each method was evaluated using Mean Reciprocal Rank (MRR), Mean Average Precision (MAP), and Normalized Discounted Cumulative Gain at rank 5 (NDCG@5). Table 3 presents the comparative results.

**Table 3.** Performance comparison between baselines and the proposed method. The asterisk ($*$) indicates a statistically significant difference ($p < 0.01$) compared to the strongest Baseline 2 (Popularity Only) based on the Wilcoxon Signed-Rank Test. Bold values indicate the best performance for each metric.

| Method | MRR | MAP | NDCG@5 |
|---|---|---|---|
| Baseline 1 (Accessibility Only) | 0.237 | 0.824 | **0.827** |
| Baseline 2 (Popularity Only) | **0.424** | **0.834** | 0.794 |
| **Proposed Method ($Score_{final}$)** | 0.397 | 0.829 | 0.814 $*$ |

The results in Table 3 reveal a critical trade-off between logistical feasibility and content interest, illustrating the strength of the proposed integrated approach as a balanced solution.

Baseline 2 (Popularity Only) achieved the highest MRR (0.424) and MAP (0.834). It confirms that the primary motivation for pilgrimage is indeed the content; users' "number one choice" is almost always a popular site. However, its NDCG@5 dropped to 0.794, the lowest among the three. It suggests that relying solely on popularity may lead users to consider locations that are logistically impractical for a top-5 itinerary, potentially reducing satisfaction in real-world planning.

Conversely, Baseline 1 (Accessibility Only) achieved the highest NDCG@5 (0.827) but a critically low MRR (0.237). It suggests that while users value convenience when filling out their top-5 list, prioritizing distance alone fails to capture the users' favourite destinations (the "top 1"), resulting in a recommendation list that is "easy to go to" but lacks "emotional value".

The Proposed Method successfully bridges this gap. It maintains a high MRR (0.397) and MAP (0.829), comparable to the Popularity baseline, while achieving a strong NDCG@5 (0.814), significantly better than the Popularity baseline. Unlike single-metric approaches that have severe limitations, such as low NDCG for popularity and low MRR for accessibility, the integrated model offers a strong and practical compromise. It ensures that recommendations are not only emotionally appealing but also logistically feasible.

Regarding **Statistical Significance and Practical Implications**, we first conducted a Wilcoxon Signed-Rank Test ($N = 20$) to ensure that the performance improvements shown in Table 3 are not due to random variance. The analysis confirms that the Proposed Method

https://doi.org/10.3390/electronics15020419

significantly outperforms the strongest content competitor, Baseline 2 (Popularity Only), with a $p$-value of $\approx 0.0094$ ($p < 0.01$).

Given the limited sample size ($N = 20$), the statistical tests in this study are intended to provide supportive evidence rather than definitive proof of population-level effects. The reported $p$-values should therefore be interpreted with caution, particularly for comparisons near conventional significance thresholds.

Beyond statistical evidence, we also consider the practical value of these metrics. While a marginal improvement in NDCG (e.g., $+0.02$) may appear modest numerically, it translates to a tangible enhancement in user experience. In top-$k$ recommendations, this improvement often indicates that highly relevant and accessible locations are successfully promoted to the top positions (e.g., moving from Rank 3 to Rank 1). For users, this minimizes the cognitive load of filtering out "popular but unreachable" destinations, thereby allowing them to discover feasible pilgrimage plans more efficiently.

### *4.4. Ablation Study Results*

To rigorously quantify the contribution of each factor, we conducted an ablation study by systematically removing components from the full model. The quantitative results are presented in Table 4 and visualized in Figure 4.

**Table 4.** Ablation study results for different score combinations. The asterisk ($*$) indicates a statistically significant difference in ranking distributions compared to Model C (w/o Seasonality) based on the Wilcoxon Signed-Rank Test ($p < 0.05$). Statistical significance reflects differences in ranking behavior, rather than superiority in rank-based accuracy metrics.

| Condition | Ranking (ID Order) | MRR | MAP | NDCG@5 |
|---|---|---|---|---|
| Model A (Full: All Factors) | E, G, M, B, A, D, J, I, L, H, C, F, K | 0.397 | 0.829 | 0.814 $*$ |
| Model B (w/o Accessibility) | E, L, G, M, D, B, J, A, I, C, K, F, H | 0.400 | 0.819 | 0.784 |
| Model C (w/o Seasonality) | E, G, B, A, M, D, L, J, C, I, H, F, K | 0.403 | 0.855 | 0.836 |
| Model D (w/o Popularity) | H, G, E, M, A, B, I, J, F, C, D, K, L | 0.250 | 0.807 | 0.751 |

https://doi.org/10.3390/electronics15020419

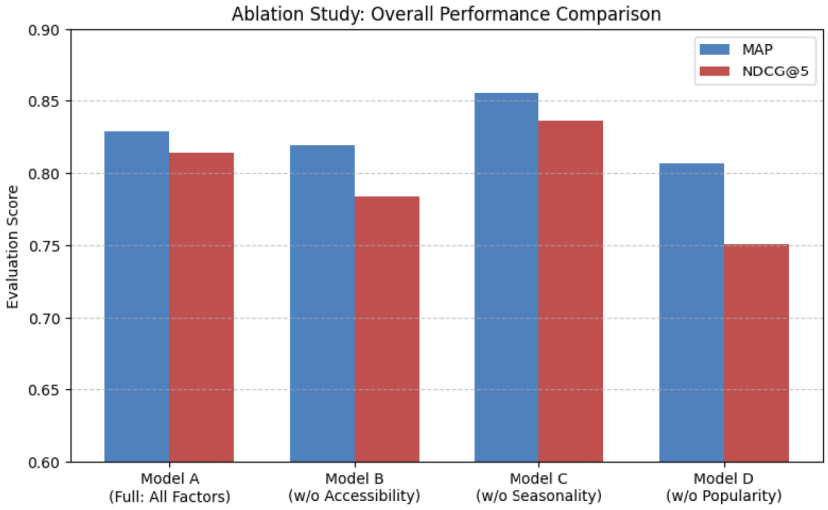



**Figure 4.** Ablation Study: Overall Performance Comparison. The chart displays MAP and NDCG@5 scores across four model variants. Note the drop in performance for Model D and Model B compared to Model C in terms of MAP and NDCG@5.

#### 4.4.1. Impact of Popularity and Accessibility

As clearly illustrated in Figure 4, the performance varies significantly across different model variants.

- **Impact of Popularity (Model D):** The rightmost bars in Figure 4 show the lowest performance across all metrics (NDCG@5: 0.751). This visual drop confirms that popularity acts as the foundational filter for user interest.
- **Impact of Accessibility (Model B):** Comparing Model B with the Full Model (Model A), we observe a noticeable decline in the NDCG score (red bar), indicating that ignoring logistical constraints negatively impacts the quality of the top-5 recommendations.

#### 4.4.2. Analysis of Seasonality Impact (Model C)

Interestingly, Model C (w/o Seasonality) achieved the highest metrics (MAP: 0.855). We attribute this phenomenon to the discrepancy between the user evaluation task and the model's design objective.

In the user study, the ranking task primarily elicited "static" content preferences—ranking locations based on their interest in the anime works and the intrinsic fame of the sites—without explicitly considering current temporal constraints. Conversely, the Full Model incorporates "dynamic" feasibility, penalizing locations with high seasonal gaps to ensure realistic travel planning. When the model downgrades a popular but off-season location, it diverges from the user's static preference ranking, resulting in a mathematical decrease in MAP.

Therefore, the lower MAP in the Full Model reflects a necessary trade-off between maximizing immediate user interest match (Static Relevance) and ensuring actionable travel timing (Dynamic Feasibility). This observation does not imply that excluding seasonality yields a superior recommendation model, but rather highlights a mismatch between static relevance-oriented evaluation metrics and planning-oriented model objectives.

It should be noted that the terms "feasibility" and "actionability" in this context refer strictly to model-imposed temporal constraints reflecting seasonal alignment, rather than empirically validated user decision outcomes or actual travel behavior.

https://doi.org/10.3390/electronics15020419

#### 4.4.3. Statistical Validation of Components

We further applied the Wilcoxon Signed-Rank Test to the ablation results in Table 4 to verify the contribution of individual modules. The comparison between the Full Model (Model A) and the strongest ablation variant, Model C (w/o Seasonality), reveals a statistically significant difference (Wilcoxon Signed-Rank Test, $p \approx 0.041$, $p < 0.05$). Although Model C achieved slightly higher raw metrics, the statistically significant difference indicates that the Seasonality module induces systematic changes in ranking behavior by filtering out temporally misaligned (but popular) options, rather than acting as a passive component.

### *4.5. Parameter Sensitivity Analysis*

To justify the selection of weights ($\alpha = 0.3, \beta = 0.2, \gamma = 0.5$) used in our proposed method, we conducted a sensitivity analysis to observe how variations in the importance of "Anime Popularity" ($\gamma$) affect system performance. Since popularity is the dominant factor in user attention, we varied $\gamma$ from 0.1 to 0.9 in increments of 0.2. To satisfy the constraint $\alpha + \beta + \gamma = 1.0$, the remaining weight $(1 - \gamma)$ was distributed to Accessibility ($\alpha$) and Seasonality ($\beta$) while maintaining the ratio of Accessibility to Seasonality constant at $\alpha : \beta = 3 : 2$. This experimental design isolates the effects of the trade-off between "content power" (Popularity) and "contextual constraints" (Accessibility and Seasonality).

Figure 5 illustrates the MAP and NDCG@5 scores across different $\gamma$ values. The results reveal a clear trend: ranking performance improves significantly as $\gamma$ increases from 0.1 to 0.5, reflecting the strong correlation between anime popularity and user interest. However, beyond $\gamma = 0.5$, the performance gains exhibit diminishing returns.

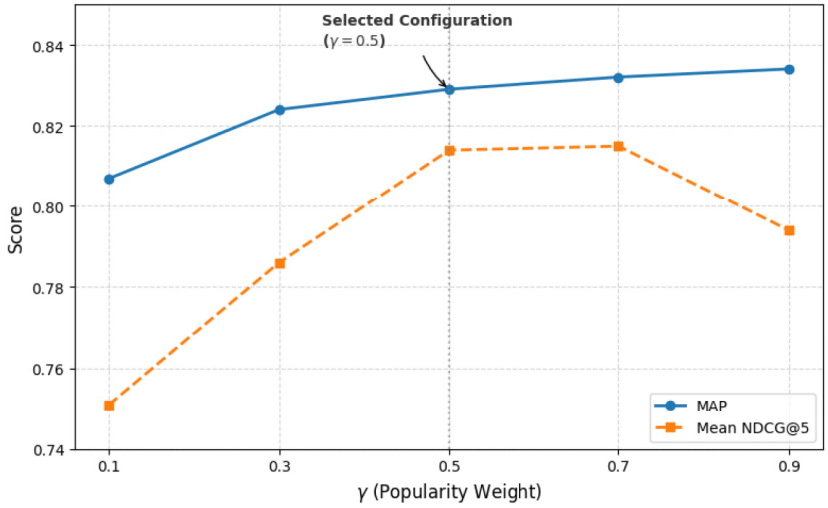



**Figure 5.** Sensitivity analysis of the Popularity weight ($\gamma$) against performance metrics (MAP and NDCG@5). The performance plateaus after $\gamma = 0.5$, indicating that further increasing popularity weight yields negligible gains while severely compromising accessibility constraints.

Specifically, comparing the chosen configuration ($\gamma = 0.5$) with the numerical peak ($\gamma = 0.7$):

- **Marginal Performance Gain:** Increasing $\gamma$ to 0.7 yields only a negligible improvement in MAP (0.829 $\rightarrow$ 0.832) and NDCG@5 (0.814 $\rightarrow$ 0.815).
- **Significant Utility Loss:** However, this shift reduces the Accessibility weight ($\alpha$) drastically from 0.30 to 0.18.

Notably, at $\gamma = 0.9$, while MAP increases slightly, NDCG@5 actually drops (0.815 $\rightarrow$ 0.794), suggesting that excessive popularity weighting harms the precision of top-

https://doi.org/10.3390/electronics15020419

ranked recommendations. This indicates that overriding spatial and temporal constraints with extreme popularity weights introduces noise into the top-tier results.

For a travel planning system, accessibility is a hard constraint—a location that is too difficult to reach offers zero utility, regardless of its popularity. Therefore, considering both the "diminishing returns" in accuracy and the degradation of top-rank precision at extremes, we identify $\gamma = 0.5$ as the optimal equilibrium point. It achieves robust ranking accuracy without sacrificing the physical feasibility of the recommended itinerary.

## 5. Discussion

In this section, we interpret the experimental findings from the perspective of intelligent data analysis, focusing on the hierarchical structure of user motivation and the trade-off between ranking accuracy and experience quality.

### *5.1. Trade-Off Analysis: Why Balance Matters*

The comparison with baselines (Table 3) reveals the inherent limitations of single-metric recommendation strategies and the robustness of our integrated approach.

#### 5.1.1. The Trap of Single Metrics

- Baseline 1 (Accessibility Only) achieved the highest NDCG@5 (0.827) but a critically low MRR (0.237). It suggests that while proximity makes for an "easy" Top-5 list, it fails to capture the user's core motivation ("I want to see this anime"), resulting in a recommendation list that is logically convenient but emotionally irrelevant.
- Baseline 2 (Popularity Only) achieved the highest MRR (0.424) and MAP (0.834). It confirms that a user's "top choice" is almost always driven by content interest. However, its lower NDCG@5 (0.794) indicates that a purely popularity-driven list often includes distant locations that are impractical for a short itinerary, potentially forcing users to discard recommendations during actual planning.

#### 5.1.2. The Robustness of the Integrated Model

The Proposed Method ($Score_{\text{final}}$) successfully bridges this gap. While it slightly trails Baseline 2 in MAP (0.829 vs. 0.834), it significantly improves upon the logistical feasibility (NDCG: 0.814 vs. 0.794). Unlike the baselines, which suffer from extreme weaknesses (e.g., Baseline 1's low MRR or Baseline 2's lower feasibility), the integrated model provides a robust compromise. It ensures that the recommended sites are not only popular enough to motivate a visit but also reachable enough to be included in a realistic travel plan.

### *5.2. The Seasonality Paradox: Planning vs. Experiencing*

A counter-intuitive finding was that Model C (w/o Seasonality) achieved slightly higher ranking metrics than the Full Model (Model A: Proposed Method). As hypothesized in Section 4, this reflects the discrepancy between "Planning" and "Experiencing".

During the planning phase, as simulated in our questionnaire, users prioritize *feasibility* (Can I go?) and *motivation* (Do I want to go?) over *timing* (Is it the right season?). Consequently, seasonality adds noise to the prediction of user choices. However, for a recommendation system to be truly "context-aware," it must look beyond simple click-through rates. Recommending a winter scene in summer may satisfy a user's immediate curiosity (high ranking accuracy) but lead to a disappointing on-site experience (low satisfaction). Therefore, we argue that the Seasonal Score serves as a quality assurance mechanism. Although it does not immediately boost ranking metrics, it enhances the long-term value of the pilgrimage as a "soft constraint".

https://doi.org/10.3390/electronics15020419

### *5.3. Social and Ethical Implications*

Beyond technical performance metrics, the proposed recommendation framework holds significant implications for the sustainability of tourism and the understanding of user behavior.

Regarding the **Mitigation of Overtourism**, popular anime pilgrimage sites often face the challenge of "overtourism," where excessive visitor numbers disrupt local ecosystems. By integrating "Seasonality" and "Accessibility" alongside popularity, our system introduces a mechanism to potentially redistribute tourist traffic. For example, by identifying off-peak seasons (via Seasonality) or recommending accessible but less-visited locations (high Accessibility, moderate Popularity), the framework can guide tourists toward alternative options. This helps balance visitor flows, reducing the pressure on hotspots while promoting hidden gems.

From a sociological perspective, the "Popularity" module functions as more than just a metric of fame; it serves as a proxy for **Social Influence and Conformity**. In the context of technology adoption, users are more likely to accept recommendations that align with their lifestyle constraints (e.g., time availability and physical capability). Our system bridges the gap between the desire for social conformity (sharing collective memories) and logistical feasibility. However, it is essential to balance this motivation with ethical considerations. Since many anime scenes are inspired by actual residential areas, future iterations of the system should consider incorporating "residential privacy" constraints (e.g., flagging quiet zones) to ensure that tourism promotion does not infringe upon the daily lives of residents.

### *5.4. Limitations and Future Work*

While the proposed framework demonstrates effective multi-source fusion, we acknowledge several limitations that define the scope for future research.

#### 5.4.1. Data Reliability and Standardization Constraints

We acknowledge two primary limitations in the current data processing pipeline regarding the Cross-Platform Popularity Index (CPPI).

First, regarding **Platform Bias**, the current model employs Min-Max scaling to map diverse metric ranges into a uniform $[0, 1]$ interval. Although this method is effective for ensuring positive inputs for the weighted combination model, it is sensitive to outliers, and it does not account for varying user base distributions across different platforms (e.g., MyAnimeList vs. Bangumi). Future iterations should use Z-score standardization or robust scaling to correct for platform-specific biases and more accurately align statistical distributions.

Second, regarding **Data Reliability**, the aggregation of raw vote counts assumes that the source data is organic. Although the proposed Cross-Platform Popularity Index (CPPI) inherently mitigates single-platform anomalies (e.g., a specific site being targeted) by averaging data from four distinct user bases, it remains vulnerable to coordinated external manipulation, such as widespread "review bombing" or automated bot activity. While the current system relies on the moderation mechanisms of the source platforms, this is not a failsafe solution. Future work will necessitate the integration of an outlier detection module to identify and downweight suspicious spikes in voting patterns, ensuring that the popularity metric reflects genuine user consensus.

#### 5.4.2. Constraints of the Mathematical Models

We acknowledge three primary mathematical limitations in the current framework components:

https://doi.org/10.3390/electronics15020419

First, for the transportation accessibility module of the **Accessibility model**, we used a linear penalty function for both travel time (Equation (1)) and the number of transfers (Equation (2)). We explicitly acknowledge that this is a simplified first-order approximation. In real-world scenarios, travel impedance often exhibits nonlinear characteristics. For instance, the perceived physical and cognitive fatigue from making three transfers is likely much greater than the cumulative effect of three single transfers. Future iterations will use nonlinear scoring functions, such as exponential decay or logarithmic scaling, to more accurately simulate psychological travel effort.

Second, regarding the **Seasonality model**, while Section 5.2 highlights the value of seasonality as a "soft constraint," our current implementation relies on a discrete four-season model. However, this discrete approach may not fully capture transitional periods (edge cases). For example, early March and late May are both categorized as "Spring," yet they offer vastly different environmental experiences. To address this, future work will implement a continuous temporal model (e.g., monthly weights) and integrate real-time environmental data (e.g., temperature or vegetation indices) to handle edge cases more precisely.

Finally, concerning the **Linear Weighted Combination (LWC) model**, a significant limitation is the reliance on static weights. The current framework applies a universal weight configuration ($\alpha = 0.3, \beta = 0.2, \gamma = 0.5$) to all users, effectively treating the user base as a homogeneous group. This static approach fails to capture user heterogeneity. For example, some users prioritize convenience over popularity, while dedicated fans may endure long travel times to visit niche destinations. To address this variability, future iterations will transition to a Dynamic Weighting Framework. We plan to implement an interactive mechanism, such as preference sliders, that allows users to clearly define their priorities. Mathematically, this transforms fixed coefficients into user-specific variables ($\alpha_u, \beta_u, \gamma_u$), thereby enabling the system to generate personalized rankings tailored to individual constraints.

#### 5.4.3. Limitations of the User Study

We acknowledge two primary limitations regarding the user evaluation in this study: the statistical power related to the sample size and the demographic diversity.

First, regarding the sample size ($N = 20$) and statistical power, practical constraints limited the number of participants. While prior HCI studies [34,35] suggest that samples of this size are often sufficient to reveal consistent preference trends among domain-aware users, we recognize that the ratio of participants to the number of ranked items (13 locations) constrains the degrees of freedom. Accordingly, the statistical analyses in this study, such as the Wilcoxon Signed-Rank Test results ($p < 0.05$ and specific concordance $p$-values detailed in Tables 3 and 4), are intended to provide supportive and exploratory evidence rather than definitive population-level inference. Although the results indicate consistent ranking behavior within this specific group, they should be interpreted with caution given the modest sample size.

Second, regarding demographic generalizability, the participants were exclusively university students from Japan and China. Specifically, the cohort consisted of 12 Chinese and 8 Japanese nationals. We conducted a qualitative comparison of ranking patterns between these two subgroups and observed broadly consistent preferences, particularly for highly popular and easily accessible pilgrimage sites. However, due to the limited sample size within each subgroup, no formal statistical tests were conducted for cross-national differences. Consequently, the findings may not fully generalize to other demographics. For instance, international tourists from Western countries may prioritize iconic or "must-

https://doi.org/10.3390/electronics15020419

visit" locations over transportation convenience due to the rarity of their visits, a preference structure that may differ from our study cohort.

To address these limitations, future studies will aim to test the system with a broader user base—increasing both the sample size to robustly assess statistical significance and the diversity of participants to validate the system's effectiveness across broader cultural contexts.

## 6. Conclusions

This study established a comprehensive recommendation framework for anime pilgrimage that intelligently integrates three heterogeneous data sources: transportation accessibility, seasonal alignment, and cross-platform popularity. By synthesizing objective geospatial constraints with subjective content appreciation, the proposed system successfully addresses the complex Multi-Criteria Decision Making (MCDM) problem inherent in niche tourism planning.

This research significantly advances our preliminary work [8] by resolving its key limitations. Specifically, we refined the popularity metrics by aggregating data from four global platforms to reduce bias, and we rigorously validated the model's effectiveness through an expanded user study and a comprehensive ablation analysis.

The experimental results demonstrate that the proposed integrated model achieves a MAP of 0.829 and an NDCG@5 of 0.814, significantly outperforming single-metric baselines. The ablation study provided critical insights into user behavior: while "Anime Popularity" and "Transportation Accessibility" serve as the primary drivers for decision-making during the planning phase, "Seasonal Alignment" functions as a qualitative enhancer that ensures immersive experiences. This confirms that the weight configuration employed in our model effectively balances the trade-off between user interest and logistical feasibility.

Although the current framework has demonstrated strong predictive capability, avenues for future research remain to meet higher technical standards. First, regarding **Mathematical Optimization**, our goal is to transition from the current linear weighting system to nonlinear adaptive models that more effectively capture complex user preferences. Second, regarding **Data Validation**, future iterations will integrate automated outlier detection to ensure the reliability of social media metrics against manipulation. Finally, to further personalize the experience, we plan to integrate Large Language Models (LLMs) to analyze unstructured reviews. This will enable the system to differentiate subtle preferences, such as "architectural backgrounds" and "character-focused scenes".

In summary, this research contributes a validated, scalable data analysis framework to the field of smart tourism. It bridges the gap between the virtual world of anime and the physical world of travel, providing a scientific foundation for supporting the growing global phenomenon of content tourism.

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

**Funding:** This research was partially supported by JSPS KAKENHI Grant Numbers JP21K17862 and JP25K15354.

**Institutional Review Board Statement:** Ethical review and approval were waived for this study due to its nature as an anonymous user questionnaire survey, which does not involve personal data collection or invasive procedures, in accordance with the "Ethical Guidelines for Medical and Biological Research Involving Human Subjects" established by the Japanese Ministry of Health,

Labour, and Welfare (MHLW) and the "Regulations for General Research Involving Human Subjects" at Yamaguchi University.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The datasets used in this study are openly available as follows: The Bangumi Anime Database, available at https://bangumi.tv/ (accessed on 14 February 2025); the AniList Anime Dataset, available at https://anilist.co/ (accessed on 14 February 2025); the MyAnimeList Database, available at https://myanimelist.net/ (accessed on 14 February 2025); and the Filmarks Anime Review Dataset, available at https://filmarks.com/ (accessed on 14 February 2025). Further inquiries can be directed to the corresponding author.

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

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| CPPI | Cross-platform popularity index |
| LBS | Location-Based Services |
| LLMs | Large language models |
| LWC | Linear weighted combination |
| MAP | Mean average precision |
| MCDM | Multi-criteria decision making |
| MRR | Mean reciprocal rank |
| NDCG | Normalized discounted cumulative gain |
| POI | Point-of-Interest |

## References

1. Oishi, G. Introduction to the history of ANIME tourism : Newly revised edition (in Japanese). *Bull. Toyama Prefect. Univ.* **2020**, *30*, 25–35.
2. Wikipedia Contributors. Seichi Junrei. 2025. Available online: https://en.wikipedia.org/wiki/Seichi_junrei (accessed on 12 November 2025).
3. Squire, S.J. The cultural values of literary tourism. *Ann. Tour. Res.* **1994**, *21*, 103–120. [CrossRef]
4. Jichitai Works WEB. Revitalizing the Local Area Through Anime Pilgrimages! How to Revitalize the Town Based on Case Studies (In Japanese). 2025. Available online: https://jichitai.works/articles/2944 (accessed on 12 November 2025).
5. Japan National Tourism Organization (JNTO). Inbound Visitor Statistics. 2025. Available online: https://statistics.jnto.go.jp/en/graph/#monthly (accessed on 26 February 2025).
6. Development Bank of Japan. *Content and Regional Revitalization—100 Years of Japanese Anime, Focusing on Pilgrimages to Holy Lands*; Technical Report; Development Bank of Japan: Tokyo, Japan, 2017. (In Japanese)
7. Nagata, H.; Makinojo, Y. 7 Successful Examples of "Anime x Town Revitalization": The History of Pilgrimages and How to Revitalize Local Communities (In Japanese). 2025. Available online: https://lumiarch.ntt-east.co.jp/articles/202509_anime/ (accessed on 12 November 2025).
8. Zhou, Y.; Wang, Y. Tourist Spot Recommendation for Anime Pilgrimage Considering Anime Features and Transportation Accessibility. In Proceedings of the 2025 IEEE 14th Global Conference on Consumer Electronics (GCCE), Osaka, Japan, 23–26 September 2025; pp. 557–561. [CrossRef]
9. Soni, B.; Thakuria, D.; Nath, N.; Das, N.; Boro, B. RikoNet: A Novel Anime Recommendation Engine. *arXiv* **2021**. [CrossRef]
10. Javaji, S.R.; Sarode, K. Hybrid Recommendation System using Graph Neural Network and BERT Embeddings. *arXiv* **2023**. [CrossRef]
11. Ye, M.; Yin, P.; Lee, W.C.; Lee, D.L. Exploiting geographical influence for collaborative point-of-interest recommendation. In Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '11), Beijing, China, 24–28 July 2011; pp. 325–334. [CrossRef]
12. Bao, J.; Zheng, Y.; Wilkie, D.; Mokbel, M. Recommendations in location-based social networks: A survey. *Geoinformatica* **2015**, *19*, 525–565. [CrossRef]
13. Yuan, Q.; Cong, G.; Ma, Z.; Sun, A.; Thalmann, N.M. Time-aware point-of-interest recommendation. In Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '13), Dublin, Ireland, 28 July–1 August 2013; pp. 363–372. [CrossRef]

14. Lian, D.; Zhao, C.; Xie, X.; Sun, G.; Chen, E.; Rui, Y. GeoMF: Joint geographical modeling and matrix factorization for point-of-interest recommendation. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '14), New York, NY, USA, 24–27 August 2014; pp. 831–840. [CrossRef]
15. Alnogaithan, O.; Algazlan, S.; Aljuraiban, A.; Shargabi, A.A. Tourism Recommendation System Based on User Reviews. In Proceedings of the 2019 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), Sakhir, Bahrain, 22–23 September 2019; pp. 1–5. [CrossRef]
16. Ma, W.; Wang, Y. A Sentiment-Enhanced Recommendation System for Anime Pilgrimage Tours. In Proceedings of the 2024 IEEE 13th Global Conference on Consumer Electronics (GCCE), Kitakyushu, Japan, 29 October–1 November 2024; pp. 352–356. [CrossRef]
17. Kuang, H.; Chang, L.; Bin, C.; Gu, T. Review of a Context-Aware Travel Recommendation System. *CAAI Trans. Intell. Syst.* **2019**, *14*, 611–618.
18. Sha, Q.; Ma, L.; Chen, N. A Smart Tourism Resource Information Management Platform Based on Multi-Source Data Fusion. *Int. J. High Speed Electron. Syst.* **2025**, *34*, 2540134. [CrossRef]
19. Sudarshan, K.; Bhagyalakshmi, M.V. Seasonal Tourist Recommendation Portal. *Int. J. Sci. Res. (IJSDR)* **2023**, *8*, 249–256.
20. Zheng, R.; Zhen, S.; Cai, A.; Li, S.; Liu, H. Anime tourists traveling to Japan: Pilgrimage behaviors patterns and the formation of homologous emotions. *Front. Psychol.* **2025**, *16*, 1637114 . [CrossRef] [PubMed]
21. Putra, I.W.K.A.; Priatdana, G.Y.A. Implementation of the TOPSIS Method for a Decision Support System in Recommending Tourist Destinations in Tabanan. *J. Sist. Inf. Dan Komput. Terap. Indones. (JSIKTI)* **2025**, *6*, 460–469.
22. Buasri, N.; Sangpradid, S. Enhancing Tourist Experience: A Fuzzy TOPSIS Approach to Destination Recommendation Systems. *GeoJournal Tour. Geosites* **2025**, *52*, 15–24. [CrossRef]
23. NAVITIME Japan. NAVITIME Route (Totalnavi). 2025 Available online: https://rapidapi.com/navitimejapan-navitimejapan/api/navitime-route-totalnavi (accessed on 14 February 2025).
24. Google Maps Platform. Geocoding API Overview. 2025. Available online: https://developers.google.com/maps/documentation/geocoding/overview (accessed on 14 February 2025).
25. Wardman, M. A review of British evidence on time and service quality valuations. *Transp. Res. Part E Logist. Transp. Rev.* **2001**, *37*, 107–128. [CrossRef]
26. Guo, Z.; Wilson, N.H. Assessing the cost of transfer inconvenience in public transport systems: A case study of the London Underground. *Transp. Res. Part A Policy Pract.* **2011**, *45*, 91–104. [CrossRef]
27. Schakenbos, R.; Paix, L.L.; Nijenstein, S.; Geurs, K.T. Valuation of a transfer in a multimodal public transport trip. *Transp. Policy* **2016**, *46*, 72–81. [CrossRef]
28. Wikipedia Contributors. Bangumi Fangzu Jihua (In Chinese). 2024. Available online: https://zh.wikipedia.org/zh-cn/Bangumi_ (accessed on 14 February 2025).
29. Mojimoon. Bangumi Anime Ranking Analysis. 2023. Available online: https://github.com/mojimoon/bangumi-anime-ranking (accessed on 14 February 2025).
30. Manami-Project. Anime-Offline-Database. 2025. Available online: https://github.com/manami-project/anime-offline-database (accessed on 14 February 2025).
31. Wikipedia Contributors. MyAnimeList. 2025. Available online: https://en.wikipedia.org/wiki/MyAnimeList (accessed on 14 February 2025).
32. Wikipedia Contributors. Filmarks (In Japanese). 2025. Available online: https://ja.wikipedia.org/wiki/Filmarks (accessed on 14 February 2025).
33. DOCOMO ANIME STORE Inc. Anime Viewing Numbers (Overall) Ranking (In Japanese). 2025. Available online: https://animestore.docomo.ne.jp/animestore/CR/CR00000014?ranking_type=views&period=all (accessed on 14 February 2025).
34. Nielsen, J. *Why You Only Need to Test with 5 Users*; Nielsen Norman Group: Fremont, CA, USA, 2000.
35. Faulkner, L. Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behav. Res. Methods Instruments Comput.* **2003**, *35*, 379–383. [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.

*electronics*

MDPI

*Article*

# A Sequence Prediction Algorithm Integrating Knowledge Graph Embedding and Dynamic Evolution Process

**Jinbo Qiu [1,2], Delong Cui [1,3,*], Zhiping Peng [1,4], Qirui Li [1,3] and Jieguang He [1]**

[1] College of Electronic Information Engineer, Guangdong University of Petrochemical Technology, Maoming 525000, China
[2] The Major of Computer and Software Engineering, Mongolian National University, Ulaanbaatar 15141, Mongolia
[3] Guangdong Provincial Key Laboratory of Petrochemical Equipment Fault Diagnosis, Maoming 525000, China
[4] Information Engineering College, Jiangmen Polytechnic, Jiangmen 529030, China
* Correspondence: delongcui@gdupt.edu.cn

**Abstract**

Sequence prediction is widely applied and has significant theoretical and practical application value in fields such as meteorology and medicine. Traditional models, such as LSTM(Long Short-Term Memory) and GRU(Gated Recurrent Unit), may perform better than this model when dealing with short-term dependencies, but their performance may decline on long sequences and complex data, especially in cases where sequence fluctuations are significant. However, the Transformer requires a large amount of computing resources (parallel computing) when dealing with long sequences. Aiming to solve the problems existing in sequence prediction models, such as insufficient modeling ability of long sequence dependencies, insufficient interpretability, and low efficiency of multi-element heterogeneous information fusion, this study embeds sequential data into the knowledge graph, enabling the model to associate context information when processing complex data and providing more reasonable decision support for the prediction results. Given the historical sequence and the predicted future sequence, three groups of sequence lengths were set in the experiment. And MAE (Mean Absolute Error)and MSE (Mean Square Error) are used as indicators for sequence prediction. In sequence prediction, dynamic evolution is conducive to enhancing the ability of the prediction model to capture the changing patterns of the current time series data and significantly improving the reliability of the prediction results. Experiments were conducted using five datasets from different application fields to verify the effectiveness of the prediction model. The experimental results show that based on the randomization of the prediction time step, the prediction model proposed in this study significantly improves the expression performance of stationary sequences. It has addressed the shortcomings of these traditional methods, such as maintaining good performance in the case of short sequences with large fluctuations.

**Keywords:** attention mechanism; dynamic evolution; key hyperparameters; knowledge graph embedding; sequence prediction

## 1. Introduction

Sequence prediction refers to using data from a certain period in the past to predict information for a certain period in the future. It includes continuous prediction (numerical prediction and range estimation) and discrete prediction (event prediction). It has extremely high commercial value. Its main task is to predict the future value of a certain indicator

based on its historical data. Sequence prediction has a wide range of applications, such as meteorological prediction [1], 2.5-μm Particulate Matter prediction [2], traffic congestion prediction [3], and so forth, providing technical support for many information application scenarios in society.

The long short-term memory network (LSTM) and the gated recurrent unit (GRU) models have enhanced the ability of sequence prediction models to handle long sequence data and capture long-term dependencies from a new technical perspective. For example, Hidasi et al. adopted the idea of the GRU model in the recommendation system [4], that is, using the GRU to model the sequence and capture the relationship of behavioral changes in the long sequence.

The application of the attention mechanism [5] has led to significant technological advancements in information processing methods, particularly in terms of long-distance dependent capture. Wen et al. proposed using a double-layer LSTM as the encoder and the decoder [6]. An attention mechanism was introduced between the encoder and the decoder, thereby enabling the model to capture the key information in the input sequence more accurately. The Transformer proposes the self-attention mechanism and parallel processing capability [7] to address the problem of long-distance dependencies and is now widely used in sequence prediction tasks. The SASRec (Self-Attentive Sequential Recommendation) model adopts the Transformer encoder as its backbone for sequence prediction [8]. This method alleviates the overly dependence of the sequence model on the current input variable by obtaining the global information of the sequence.

The sequence prediction task is diverse, which is mainly reflected in the prediction of multiple related variables. Wang et al. proposed that sequence data should account for both target features and auxiliary information, which exhibit extremely complex spatial dependencies. To address these issues, Wang et al. proposed the AI-DTN (Digital Twin Network) model [9]. This model designed a GRU (AI-GRU) acting on auxiliary information. Three different gating units were added to the AI-GRU for coordinating the calculation of target features and auxiliary information. The AI-DTN model conducted autonomous reinforcement learning on the influence of different variables on the results by effectively integrating sequence information, target features, and auxiliary information.

The GNN architecture has been widely applied in the in-depth research of graph data structures, such as graph convolutional network [10] (GCN) and graph attention network [11] (GAT). GCN has the same function as the basic deep learning model CNN and is used for feature extraction. Graph data, when processed by a GCN, can effectively capture the characteristics of nodes and their neighbors, thus enhancing the model's understanding of the graph structure. In a GCN, the weight of message passing depends only on the degree of nodes. In contrast, GAT adopts an attention mechanism to iteratively update the weights of message passing in the form of attention scores learned with the model parameters. GAT can effectively and dynamically aggregate information from neighboring nodes. We propose another idea for the research on graph sequence data augmentation. By estimating the emission probability, nodes within a sequence can be connected to nodes outside the sequence, thereby generating a new sample sequence and achieving graph sequence data augmentation [12].

Knowledge graphs are also a type of graphs. They establish different semantic connections among sample data in the form of triples (head entity, relationship, and tail entity). They can effectively capture complex data and ultimately extract the features of knowledge interaction. Knowledge graphs are widely applied in various application scenarios, such as question-answering systems [13], recommendation systems [14], and so forth. Significant progress has also been made in the exploration of methods integrating sequence prediction with knowledge graphs. This method uses the triplet information extracted from the knowl-

edge graph and hence enables the model to better capture the user's interest characteristics and their evolution, thereby achieving more precise personalized recommendations in sequence prediction.

Sequence prediction, as the core task of time series data analysis, has significant application value in scenarios such as meteorology, medical care, and recommendation systems. In recent years, dynamic evolution modeling has become a key direction for improving the performance of sequence prediction, with the core lying in capturing the dynamic change patterns of the system over time. Most existing studies are based on deep learning frameworks. For instance, DIEN tracks user behavior changes through an interest evolution layer, DNE and JODIE, respectively, depict structural evolution from the perspectives of dynamic network embedding and node trajectory modeling, while DySAT captures evolution patterns in dynamic graph structures with the help of graph attention networks and temporal snapshots. Although these methods perform well in specific tasks, existing sequence prediction models still have obvious limitations: First, their modeling ability for complex sequences (high-dimensional, nonlinear, and implicit patterns) is limited, and traditional models and some depth methods find it difficult to accurately capture their intrinsic patterns; Second, there is insufficient long-term dependence modeling. The model is easily dominated by short-term dependence, and the remote associated information gradually decays during the transmission process. Thirdly, the ability to cross-integrate multiple relationships is weak. The complex dependencies among multiple variables introduce noise and affect the stability of the model. In addition, semantic enhancement methods such as graph structures, while introducing rich relational information, also bring significant computational burdens, restricting the scalability of actual systems. To address the above challenges, this paper proposes a dynamic evolution sequence prediction method integrating knowledge graph embeddings. This method employs sequence data into a dynamic knowledge graph to explicitly model the temporal correlations and semantic relationships between entities, thereby enhancing the ability to capture long-term dependencies and the interpretability of the model. Meanwhile, an efficient structural evolution mechanism should be designed to control the computational complexity while introducing external knowledge, thereby enhancing the prediction accuracy and robustness for multivariate complex sequences.

Sequence prediction, as the core task of time series data analysis, is gradually shifting from relying solely on historical value matching to exploring the complex entity relationships and evolution patterns behind the data. Against this backdrop, knowledge graph embeddings are introduced into sequence prediction, providing the model with rich semantic context. By mapping entities in the sequence (such as users, products, and symptoms) to a low-dimensional vector space, knowledge graph embeddings can explicitly encode multi-hop associations (such as class, subordination, and causality) among entities, enabling the model not only to learn temporal patterns but also to perform relational reasoning, thereby enhancing the interpretability of predictions, especially when the data is sparse or there are cold start problems.

However, static knowledge graphs cannot fully reflect the dynamics of relationships in the real world. Therefore, dynamic evolution modeling becomes a key supplement. This type of method aims to capture the structural changes in the system over time, for example, by modeling the relationships of addition, disappearance or intensity change through temporal knowledge graphs or dynamic graph neural networks. This, when combined with sequence modeling, enables the model to synchronously track the evolution of entity states and the changes in their relationship networks, thereby more accurately capturing long-term dependencies. In this process, the attention mechanism plays a core role. It can adaptively weigh the importance of different historical time steps and different associations

between entities. For instance, when predicting the next event, the model can focus on recent key events and their related entities, effectively filtering out noise.

At present, hybrid architectures that integrate these ideas have become mainstream and can mainly be classified into the following categories for comparative analysis:

GCN/ GNN-RNN hybrid model: This type of model typically uses GNN to encode the graph structure information at the current moment and then inputs the output to RNN (such as LSTM/GRU) to capture temporal dependencies. Its advantage lies in its clear structure, and RNN itself is a powerful sequence model. However, its limitations lie in the inherent sequential processing method and vanishing gradients of RNN, which make it difficult to effectively capture very long-term dependencies and result in relatively low computational efficiency.

Transformer–GNN hybrid model: This type of model utilizes the self-attention mechanism of Transformer to replace RNN for modeling temporal dependencies. Self-attention can perform parallel computing and directly capture global dependencies between any distant time steps, significantly outperforming RNNS on long sequences. GNN is responsible for extracting structural information at each time step. The combination of the two (such as GraphTransformer) is currently a cutting-edge direction, but its challenge lies in the fact that the computational complexity increases exponentially with the sequence length, and it requires meticulous design to effectively integrate the graph structure information into the attention calculation.

Specialized architectures based on dynamic evolution: In addition to the above-mentioned general hybrid models, architectures specifically designed for dynamic graphs (such as DyRep and TGN) have emerged. They treat sequential events (such as interactions) as continuous flows and update the embeddings of affected nodes each time an event occurs. This type of model seamlessly integrates sequence prediction with graph evolution, making it particularly suitable for scenarios with frequent updates. However, it places higher demands on the model's continuous-time modeling capability and computational efficiency.

In conclusion, dynamic sequence prediction based on knowledge graph embedding represents an important development direction in this field. Through comparison, it can be known that the Transformer–GNN hybrid model has the greatest potential in capturing long-range and global dependencies. However, how to balance its expressive power and computational cost, and more precisely integrate temporal and structural attention, remains a key gap and challenge in current research.

## 2. Sequence Prediction, Knowledge Graph Embedding, and Dynamic Evolution Process

### *2.1. Problem Definition of the Sequence Prediction Model*

The sequence prediction method achieves the purpose of predicting future-related trends in the scene by analyzing and modeling historical sequences. Suppose a historical sequence sample $X_{1\sim L} = \{x_1, x_2, ..., x_{L-1}, x_L\}$, where $(x \in X_{1\sim L})$ is a single sequence sample. The expression of the sequence prediction model is shown in Equation (1), where $\hat{y}_{1+L}$ represents the predicted value at the next time point, F represents the sequence prediction model, and L represents the length of the historical sequence.

$$\hat{y}_{1+L} = F(X_{1\sim L}) \tag{1}$$

Sequence prediction models are applied in various application scenarios to predict the possible outcomes of a certain event in the future. The practical application examples of Equation (1) are expounded by listing several typical application scenarios.

(1) Sequence prediction based on classification: Common scenarios include the click-through rate (CTR) prediction task [15]. CTR is one of the important indicators for measuring the benefits of a product. In the sequence prediction task, $\hat{y}_{1+L}$ has only two values: 0 and 1.0, respectively, representing that the user dislikes or likes the product.

(2) Sequence prediction based on regression and multi-step time: Common regression scenarios include temperature, humidity, and other indicators. The values of these indicators exist within a certain range [16]. A typical multi-step application scenario of time-series prediction is weather forecasting. This scenario not only takes the next time prediction point as a parameter but also simultaneously inputs multiple future time prediction points into the model for prediction purposes. If the set parameters are as $\hat{Y}_{L\sim T} = \{\hat{y}_L, \hat{y}_{L+1}, ..., \hat{y}_{L+T}\}$, then a single future point in time is denoted as $\hat{y}(\hat{y} \in \hat{Y}_{L\sim T})$. Here, T represents the length of the future time point. The expression of this type of sequence prediction model is as:

$$\hat{Y}_{L\sim T} = F(X_{1\sim L}). \tag{2}$$

### *2.2. Embedding of Knowledge Graphs*

The knowledge graph is one of the key technologies in artificial intelligence, first proposed by Google in 2012. It is a structured semantic knowledge base that represents semantic relationships between entities in the form of triples (head entity, relation, and tail entity) [17]. Compared with other graph computing approaches, the symmetrical edges of a knowledge graph are of different types. In 2013, the TransE model was proposed as the most fundamental translation distance model. Suppose the head entity vector is Head, the relation vector is Relation, and the tail entity vector is Tail. The spatial position of the TransE model is shown in Figure 1.

$$h + r = t \tag{3}$$

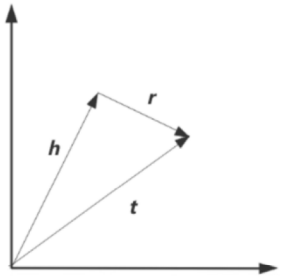



**Figure 1.** Spatial position of the TransE model.

The parameter relationship of the TransE model follows Equation (3), which indicates that the tail entity can be obtained by translating the head entity vector with the relation vector. The loss function of TransE is defined as:

$$\text{loss}_E = \max(0, \|h + r - t\|_2 - \|h' + r - t'\|_2 + m) \tag{4}$$

In the aforementioned equation, $h'$ and $t'$ are, respectively, the head entity vector and the tail entity vector obtained by negative sampling. $\|h + r - t\|_2$ is the vector modulus of positive sampling. $\|h' + r - t'\|_2$ is the vector modulus of negative sampling. The operation of negative sampling is adopted to avoid the problem of deliberately reducing the vector moduli of h, r, and t during the training process of the model. The max function is used to take the maximum value so as to prevent the loss value from being negative. m is a hyperparameter used to prevent phenomena such as overfitting in the model.

The training performance of the TransE model is poor in handling one-to-many, many-to-one, many-to-many, and reflexive relationships. To address these issues, TransH proposed a new model in 2014 [18]. The spatial position of the TransE model is shown in Figure 2. The core idea is to define a hyperplane $w_r$ for each relation, with $h_\perp$ and $t_\perp$ as the projections of h and t on $w_r$. The loss function of TransH can be defined as:

$$h_\perp = h - w_r^T \times h \times w_r \tag{5}$$

$$t_\perp = t - w_r^T \times t \times w_r \tag{6}$$

$$\text{loss}_E = \max(0, \|h_\perp + r - t_\perp\|_2 - \|h'_\perp + r - t'_\perp\|_2 + m) \tag{7}$$

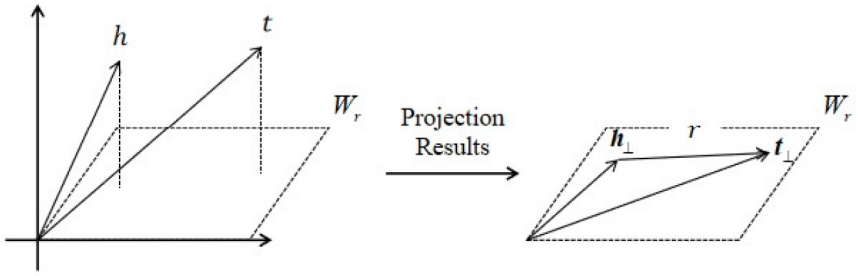



**Figure 2.** Spatial position of the TransH model.

In Equations (5) and (6), the $w_r$ vector is the normal vector of the hyperplane $w_r$. The effects of each parameter in Equation (7) are consistent with those in Equation (4). The difference between the TransH and TransR models is that the TransH model maps entities to the hyperplane of relations whereas the TransR model maps entities to the vector space of relations [19]. The schematic diagrams of the two models are shown in Figure 3.

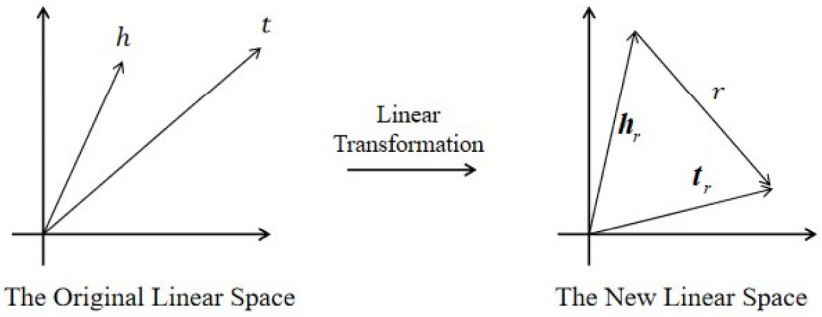



**Figure 3.** Spatial mapping diagrams of the TransH and TransR models.

TransE is suitable for simple relationships, but its ability to model complex ones is limited. After investigation, it is found that RotatE and ConvE do perform well in the knowledge graph embedding task. Considering the complexity and reproduction difficulty of the current knowledge graph embedding models, TransR, which has low computational complexity and high interpretability, was chosen.

Suppose the dimension of the r vector is d. Multiply the points h and t by a k × d matrix. The head entity hr and the tail entity tr mapped to the relation vector space can be obtained. The loss function of TransH is expressed as:

$$h_r = h \times M_r \tag{8}$$

$$t_r = t \times M_r \tag{9}$$

$$\text{loss}_R = \max(0, \|h_r + r - t_r\|_2 - \|h'_r + r - t'_r\|_2 + m) \tag{10}$$

The entity is the corresponding feature vector, precisely the vector embedded through Embedding, allowing the model to deeply explore the potential features under that date. For instance, in the eyes of the model, July is seen as hot and humid, etc. How is the hidden state of GRU generated as a relation vector? It represents the hidden feature of the KG "relation" (triple r). Just a few specific numerical features seem like statistical methods rather than deep learning methods. Just a few specific numerical features need to be transformed by machine learning into a set of relational r feature vectors that can make the dimension of the embedding layer vector equal. Ensure that the vector addition of u + r = o is feasible, rather than having one vector with a large dimension and the other with a small dimension.

### *2.3. Dynamic Evolution Method*

Dynamic evolution refers to the changes in a system over time and is widely used in sequence prediction tasks. The DIEN (Deep Interest Evolution Network) model of the recommendation system adopts the dynamic evolution method to track the changes in users' interest behaviors [20]. The interest evolution layer of the DIEN model is shown in Figure 4. The dynamic evolution process of the DIEN model is expressed as:

$$\alpha = \text{Softmax}(x_t \times W_{\text{att}} \times d) \tag{11}$$

$$\widetilde{u}_t = \alpha \times u_t \tag{12}$$

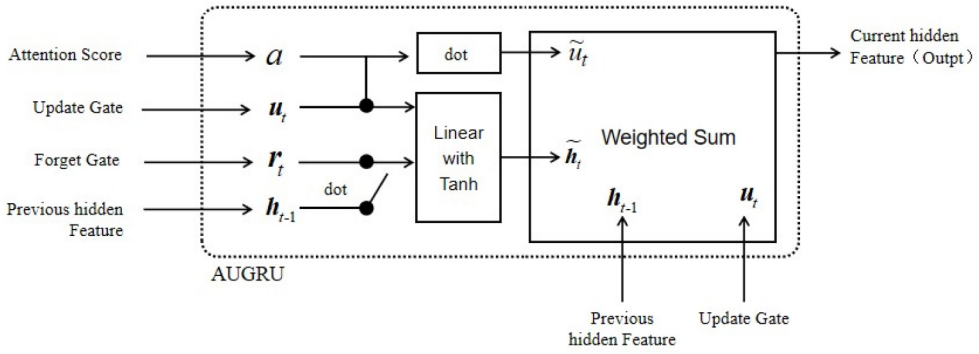



**Figure 4.** Schematic diagram of the interest evolution layer of the DIEN model.

In the aforementioned equations, the attention score $\alpha$ consists of the input $x_t$ at time t, the target item d with information enhancement, and the parameter matrix $W_{att}$. Equation (11) uses the dot product to calculate the similarity between $x_t$ and d. The higher the similarity between the two items, the greater the attention weight generated. The parameter matrix w is used for the linear transformation of x to improve the fit of the model. The attention score $\alpha$ is multiplied by the update gate $u_t$ to adjust the importance of information, ensuring that the DIEN model focuses on highly relevant input when updating the state.

The DySAT (Deep neural representation learning of dynamic self-attention networks) model applies GAT for dynamic representation learning and divides the entire time period into multiple discrete snapshots using a time window to capture the dynamic evolution of the sequence [21]. The dynamic evolution of the DySAT model is mainly guided by GAT. The message-passing mechanism of the graph attention network of the DySAT model is shown in Figure 5.

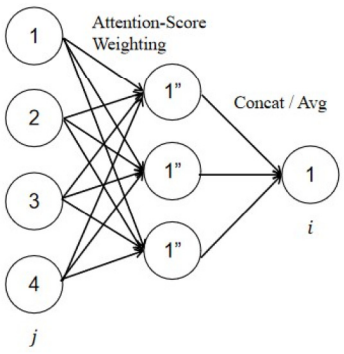



**Figure 5.** GAT information transmission process of the DySAT model.

The dynamic evolution mechanism of the DySAT model is expressed in Equations (13)–(15). Equation (13) is used to calculate the attention score $e_{ij}$ from node j to node i, where $h_i$ and $h_j$ are the input feature vectors of node i and node j, respectively. It is assumed that their dimension is F. W is the parameter matrix that the model needs to learn, with the dimension set as $F \times F'$. The symbol "||" represents the vector concatenation operation. β is also the attention vector that the model needs to learn, with the dimension of $2F'$. Finally, the result is obtained by activating the function LeakyReLU. Equation (14) is used to calculate the message-passing attention weight $a_{ij}$ from node j to node i, where $N_j$ represents the first-order neighbor set of node i. Equation (15) is used to calculate the GAT result, that is, the feature vector $h_i'$ of the new node i, where σ is an activation function and K represents the number of information transmissions. The DySAT model can aggregate key information from neighboring nodes, capture the real-time evolution of graph data during training, and enhance the reasoning and predictive ability of the model for data evolution.

$$e_{ij} = \text{LeakyReLU}(\beta^T[W \times h_i || W \times h_j]) \tag{13}$$

$$a_{ij} = \frac{exp(e_{ij})}{\sum_{k \in N_i} exp(e_{ik})} \tag{14}$$

$$h_i' = \sigma\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{j \in N_i} a_{ij}^k x^k h_j\right) \tag{15}$$

## 3. A Sequence Prediction Algorithm Integrating Knowledge Graph Embedding and Dynamic Evolution

To address issues such as insufficient modular ability, lack of interpretability, and low efficiency in fusing multi-heterogeneous information in most of the existing sequence prediction algorithms, this study proposed a sequence prediction algorithm integrating knowledge graph embedding and dynamic evolution, abbreviated as KD4SP.

### *3.1. Problem Definition of the KD4SP Algorithm*

The problem addressed by the KD4SP algorithm is defined as follows: Given a historical sequence sample $X_{1\sim L} = \{x_1, x_2, ..., x_{L-1}, X_L\}$, where $x(x \in X_{1\sim L})$ is a single sequence sample and L is the fixed sequence window size. The characteristics of this historical sequence sample can be represented as $E_{1\sim L} = \{e_1, e_2, ..., e_{L-1}, e_L\}$. Based on the aforementioned conditions, the goal is to predict the future sequence $\hat{Y}_{L+1:H} = \hat{Y}_{L+1:H} = \{\hat{y}_{L+1}, \hat{y}_{L+2}, \ldots, \hat{y}_H\}$, where H is the number of time steps for predicting the future.

### *3.2. KD4SP Algorithm Model*

The KD4SP model is shown in Figure 6. It mainly includes four modules: the knowledge graph embedding layer, the dynamic evolution layer, the feature extraction layer, and the fully connected layer.

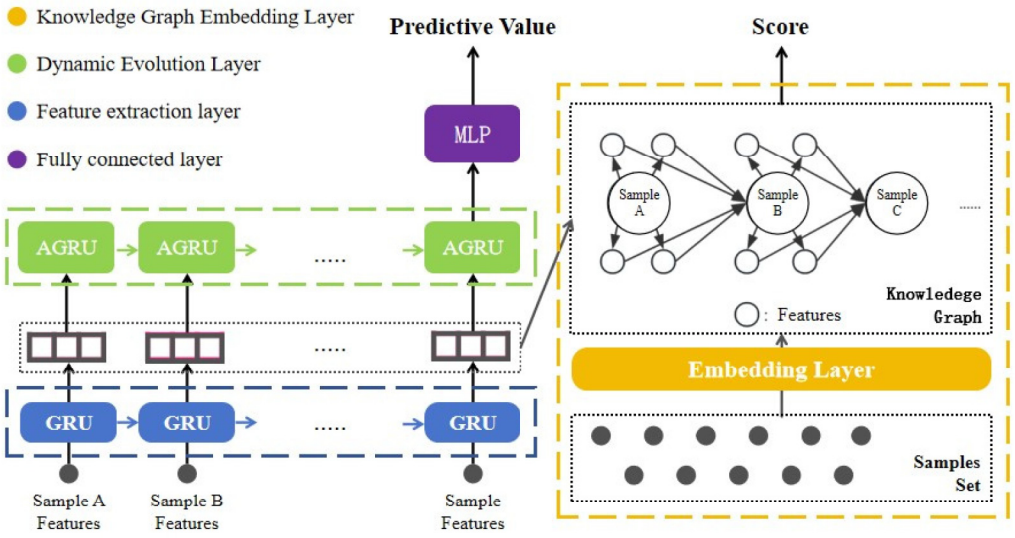



**Figure 6.** KD4SP algorithm model framework.

The KD4SP algorithm passes the sample data through the embedding layer and endows the samples with basic feature information. The features of the samples pass through the feature extraction layer, converting low-dimensional features into high-dimensional feature representations, extracting their hidden features. These features are then embedded together with the originally embedded sample data into the knowledge graph. The KD4SP algorithm inputs the hidden feature representation into the dynamic evolution layer to capture the changing trend of the sequence at this stage. Finally, the output result from the last layer of the dynamic evolution layer is then passed through a fully connected layer to obtain the predicted values. The knowledge graph embedding layer computes a scoring coefficient based on the knowledge graph embedding algorithm. The scoring coefficient mainly contributes to the loss function calculation of the model and plays a role in reinforcement learning.

Before embedding in the knowledge graph, the hidden features of the samples should be extracted to extract meaningful information and relationships. This ensures that the embedding better reflects the entities and the correlations among them, improving the performance and reasoning ability of the model. GRU is a typical sequence model. It dynamically adjusts the information flow of data through a gating mechanism, effectively capturing the long-term dependencies and key data [22]. Its structure is shown in Figure 7.

$$u_t = \text{Sigmoid}(W_u \times x_t + U_u \times h_{t-1} + b_u) \tag{16}$$

$$r_t = \text{Sigmoid}(W_r \times x_t + U_r \times h_{t-1} + b_r) \tag{17}$$

$$\tilde{h}_t = tanh(W_h \times x_t + r_t \otimes U_u \times h_{t-1} + b_h) \tag{18}$$

$$h_t = (1 - u_t) \otimes h_{t-1} + u_t \otimes \tilde{h}_t \tag{19}$$

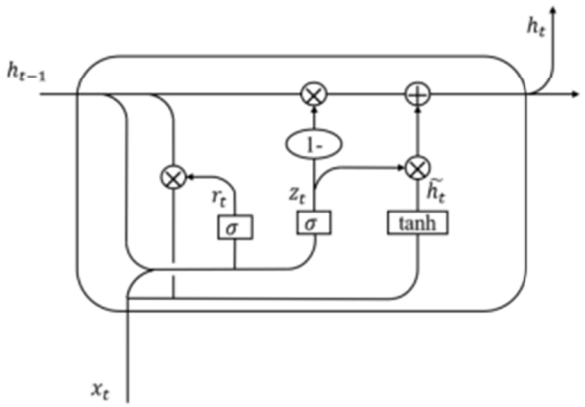



**Figure 7.** Schematic diagram of GRU model architecture.

The calculation of the GRU model is shown in Equations (16)–(19). $x_t$ represents the model input at time t; W, U, and b are the parameters to be learned during the training process, and $u_t$ and $r_t$ represent the outputs of the update gate and reset gate at time t, respectively. Both the update gate and the reset gate indicate how much hidden information from the previous moment is retained, and both are activated by the function Sigmoid. The closer the value range of $u_t$ and $r_t$ (0, 1) is to 1, the more information is retained. $\tilde{h}_t$ is the candidate hidden state, which is generated by combining the input $x_t$ at time t with the hidden state $h_{t-1}$ from the previous time step. Meanwhile, $\tilde{h}_t$ and the reset gate $r_t$ calculate the Hadamard product to control the transmission of information. The tanh activation function is used to compress the input value within the interval $(-1, 1)$ to capture the important features at the current moment. The output result $h_t$ is jointly determined by $h_{t-1}$ and $\tilde{h}_t$, and the update gate $u_t$ plays the role of weighted combination. Essentially, it is the multiplication of matrices, ensuring that each vector is in the same semantic environment after a linear transformation, while the activation function is a nonlinear transformation.

Training sequence prediction algorithms only based on sample features is insufficient. Also, the model should learn the features of the samples to improve the validity of the data. The KD4SP algorithm adopts the embedding method for sample x, and the vector is represented as its own features. Suppose the data after embedding is $x'$. As the knowledge graph embedding algorithm requires that the dimensions of entities and relations be the same, the entity features are the hidden information $h_t$ of the feature extraction layer. For the vector $x'_A$ of entity A, the vector $x'_B$ of entity B, and the feature $h_A$ of entity A, the triplet relationship is defined as:

$$x'_A + h_A = x'_B \tag{20}$$

Under the influence of its characteristics, Entity A evolves into Entity B. The KD4SP algorithm uses the TransR translation model, defining a different mapping matrix for each relation. This enables separate mapping for different relations, making the model more flexible and better able to capture complex relation features. Suppose the mapping matrix is $M_r$; the confidence score of the triplet relation can then be obtained as:

$$\text{score}(x'_A,\ h_A,\ x'_B) = \left\| M_r \times x'_A + h_A + M_r \times x'_B \right\| \tag{21}$$

The confidence score is also a parameter for measuring the loss function of the knowledge graph. The confidence value of the aforementioned triplet should be as small as possible. If the three vectors $x'_A$, $h_A$, and $x'_B$ decrease with their own moduli, score$(x'_A, h_A, x'_B)$ also gets closer to 0. Therefore, a negative sampling method can be adopted to reduce their own moduli. During this process, a random number seed is generated between 0 and 1. When the value of seed is greater than 0.5, the function score randomly selects an

entity from all entities and sets the first element member as the head pointer of the negative triplet. When the value of the seed is less than 0.5, a new sequence is generated. One entity is randomly selected from all entities, and the first element member is set as the head pointer of the positive triplet. This process is repeated in a loop until the number of positive triplets is half.

The current design models the dynamic features of samples and time steps by means of a learnable global weight vector p. Its main function is to weight the feature contributions of samples from different sequences at different time steps to capture the dynamic patterns of the data. For instance, at a certain historical moment t, if it makes a significant contribution and influences the subsequent process, then the weight value of this t in the weight vector p will be very high. The dynamic evolution layer of the KD4SP algorithm mainly comprises the AGRU (GRU with Attention) network [23]. AGRU adds an attention mechanism based on GRU. The AGRU network framework is shown in Figure 8.

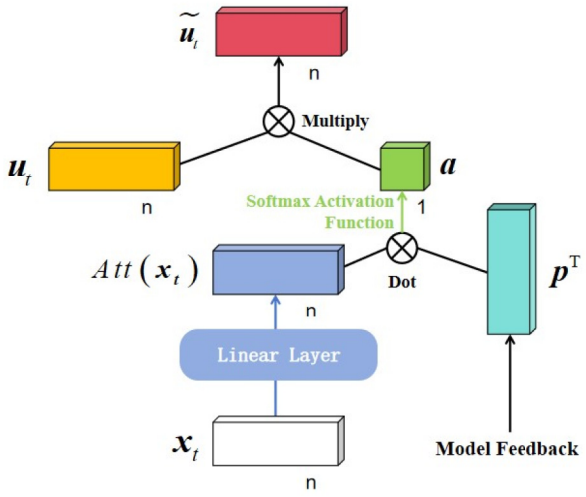



**Figure 8.** Schematic diagram of the dynamic evolution framework of the KD4SP algorithm.

The KD4SP model, by reusing the GRU state and combining it with the standard attention mechanism, although providing an efficient temporal reasoning framework, may be slightly less flexible and adaptable compared to the latest methods such as DySAT when dealing with highly dynamic and variable time series. DySAT may perform better than the KD4SP model when dealing with long-term dependencies and complex patterns, but the design of the KD4SP model places more emphasis on computational efficiency and real-time performance. The main difference between the two lies in the complexity of the dynamic attention mechanism and its adaptability to application scenarios.

Based on GRU, the calculation of the dynamic evolution process of each AGRU unit is depicted in Equations (22)–(24). $x_t$ represents the corresponding time output $h_t$ of the feature extraction layer. Equation (22) is a linear transformation of $x_t$ to improve the generalization ability of the model. $W_{att}$ and $b_{att}$ are the parameters that the model needs to learn; their dimensions are $n \times n$ and n, respectively. Equation (23) calculates the attention score at the current moment, which is formed by the dot product of $Att(x_t)$ and vector p and activated by the function Softmax. Vector p is the parameter that the model needs to learn, representing the feature weight coefficient; that is, the more important the feature, the higher the weight coefficient. The dimension of vector p is n. The dot product of vector $Att(x_t)$ and vector p is equivalent to the weighted summation effect, thus obtaining the total attention score a of all features. Then, the attention score is normalized by the activation of the function Softmax. Equation (24) resets the data of the update gate $u_t$, which is obtained by multiplying the attention score a by the update gate value $u_t$. The

updated $u_t$ determines how much of the candidate hidden information at the current moment is retained. The update gate is related to the attention score, so the model can dynamically adjust the update process of the hidden state according to the different feature weight coefficients p. As the weight coefficient p is obtained through continuous learning from historical data by the machine and through continuous attempts and adjustments, a dynamic evolution process can be achieved.

$$Att(x_t) = W_{\text{att}} \times x_t + b_{\text{att}} \tag{22}$$

$$\alpha = \text{Softmax}\left(p^{\text{T}} \times Att(x_t)\right). \tag{23}$$

$$\widetilde{u}_t = \alpha \times u_t \tag{24}$$

The fully connected layer mainly comprises a combination of linear transformations and activation functions. The activation function is not set in the last layer of the fully connected layer. Considering the complexity of sequence data, the KD4SP algorithm adopts Dice (data-dependent activation function) as the activation function in the fully connected layers of the remaining levels [10] to address the problem of vanishing gradients caused by unsmooth data. The calculation method of Dice is depicted in Equations (25) and (26). Equation (25) represents the normalization processing of batch data. $E(x)$ represents the mean of the sample, $\text{Var}(x)$ represents the variance of the sample, and $\xi$ represents the noise factor. Equation (26) is used to calculate the Dice activation function. For $\text{BN}(x)$, Sigmoid activation is adopted and expressed in the form of probability, assumed to be $p(x)$. From the perspective of normalization, $p(x)$ represents the probability that the input value x is greater than the number corresponding to a Sigmoid threshold of 0.5. For example, if an array $[-3, -2, \ldots, 7]$ exists, then the number corresponding to a Sigmoid threshold of 0.5 is 2. The difference compared with ReLU-type activation functions, such as PReLU and LeakyReLU, is that the coefficient of the neuron is weakened only when the input value x is negative, as shown in Figure 9. The expected output value of the neuron is calculated using Equation (27). From a mathematical perspective, it can be understood as: finding the expected value E. The probability $p(x)$ is multiplied by x and $\alpha x$, respectively, where $\alpha$ is a hyperparameter representing the coefficient of the neuron with weakened "negative characteristics."

$$\text{BN}(x) = \frac{x - E(X)}{\text{Var}(x) + \xi} \tag{25}$$

$$\text{Dice}(x) = x \times \text{Sigmoid}(\text{BN}(x)) + \alpha \cdot x \cdot (1 - \text{Sigmoid}(\text{BN}(x)) \tag{26}$$

$$\text{Dice}(x) = x \cdot p(x) + \alpha \cdot x \cdot (1 - p(x)) \tag{27}$$

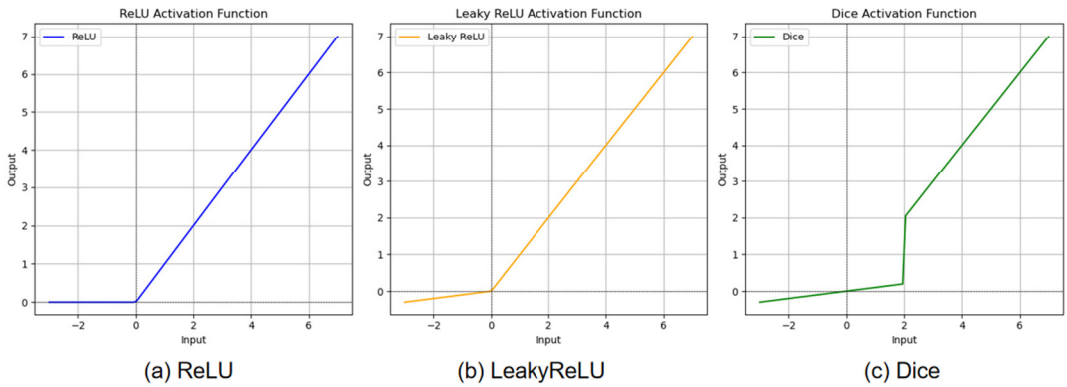



**Figure 9.** Schematic diagrams of ReLU, LeakyReLU, and Dice.

The standardized data can be de-standardized before being input into the model if other requirements exist for the output of the model. Equation (28) is used to determine anti-standardization. This is the result of standardized negation. $x_{real}$ is the de-normalized value of $x_{std}$, σ is the standard value of the original data, and μ is the mean value of the original data.

$$x_{\text{real}} = \sigma \cdot x_{\text{std}} + \mu \tag{28}$$

The loss function of the model mainly has two parts: the auxiliary loss function and the predicted loss function. The auxiliary loss function is determined by the scoring coefficient of the knowledge graph embedding algorithm. The purpose of introducing the auxiliary loss function is to improve the training effect of the mode, thereby enhancing the learning ability of the model. The predicted loss function measures the gap between the predicted and true values. The model continuously learns and adjusts its own parameters based on this measurement value to achieve the best effect.

Suppose loss is the total loss function of the model. $loss_{aux}$ and $loss_{tar}$ represent the auxiliary and predicted loss functions of the model, respectively. $\gamma$ is the weight proportion of the auxiliary loss function. The total loss function can be calculated using Equation (29). Deviations are inevitable during the training of the model. Therefore, an auxiliary loss function needs to be introduced to enhance the model's learning of the knowledge graph embedding algorithm.

$$\text{loss} = \gamma \cdot \text{loss}_{\text{aux}} + \text{loss}_{\text{tar}} \tag{29}$$

The loss function of the knowledge graph embedding algorithm is influenced by the positive and negative sampling confidence scores [24]. $x'_A$, $h_A$, and $x'_B$ are the head entity, relation, and tail entity of positive sampling. Assuming that the negative sampling is $x'_A$, $h_A$, and $x'_B$, the auxiliary loss function can be calculated using Equations (30) and (31). The auxiliary loss value is calculated using Equation (30). m is a hyperparameter used to prevent phenomena such as overfitting of the model. Equation (31). ensures that the auxiliary loss value is non-negative.

$$\text{loss}'_{\text{aux}} = \text{score}(x'_A, h_A, x'_B) - \text{score}(\hat{x}'_A, \hat{h}_A, \hat{x}'_B) + m \tag{30}$$

$$\text{loss}_{\text{aux}} = \begin{cases} \text{loss}'_{\text{aux}} & \text{if loss}'_{\text{aux}} \geq 0 \\ 0 & \text{if loss}'_{\text{aux}} < 0 \end{cases} \tag{31}$$

$$\text{MSE} = \frac{1}{K}\sum_{k=1}^{K}(y_k - \hat{y}_k)^2 \tag{32}$$

This study adopted the mean square error (MSE) as the prediction loss function. This was mainly because the MSE increased the penalty factor of the error. MSE is extremely sensitive to error, causing the model to place greater emphasis on data with large deviations from the true value during learning. The MSE can be calculated using Equation (28). k represents the sample size, and $y_k$ and $\hat{y}_k$ represent the true and predicted values, respectively. The smaller the calculation result of Equation (32), the better the performance of the model fitting.

### *3.3. Construction of Knowledge Graph*

Take the dataset Electricity as an example. The process of extracting entities and relationships from structured tables and storing them in the Neo4j graph database (Neo4j Community Edition 3.5.30 ) in the form of triples. The basic information of the dataset "Electricity" consists of date, electricity consumption, electricity generation, atomic energy, wind energy and other information.

The specific steps for constructing the knowledge graph of the dataset Electricity are as follows: First, clearly define the information form of the triples. In the sequence prediction task of this section, dates and feature values are used as nodes. Construct two types of triples: (current date, feature, current feature value) and (current feature value, influence, next day's date). Secondly, import the neo4j graph database. Based on the above form, the logic for creating head and tail nodes and relationships is written using the py2neo library(python 3.8). Traverse each row of data in the table, create nodes for samples and their features, establish relationships between each sample node and its feature nodes, and establish an influence relationship between the previous feature node and the current sample node. A detailed description of the pseudo-code algorithm is provided in Appendix A. The KD4SP model maps the features at each time point to entities, with u + r = o as the training objective. That is to say, entity u evolves into o under the influence of relation r. Therefore, it is an end-to-end learning process.

This algorithm implements the following core logic: Construction of temporal relationship: The feature nodes of the time step on the last_feature_node_group are cached to ensure the correct establishment of the connection relationship on the time series. Graph structure pattern: Sample node → feature node: HAS_FEATURE relationship, feature node → next time step sample node: TEMPORAL_FLOW relationship. Data flow control: After processing all the features of a sample each time, clear the cache of the previous time step and prepare to receive new feature nodes. This design is suitable for constructing feature graphs with temporal dependencies and is particularly suitable for graph representation learning scenarios of time series data.

The KD4SP framework mainly consists of four modules: the knowledge graph embedding layer, the feature extraction layer, the dynamic evolution layer, and the fully connected layer. Among them, the knowledge graph embedding layer helps the model capture the context information of sequential data, thereby enhancing the model's understanding and reasoning ability of the data. In addition, this paper also proposes a negative sampling method to solve the problem that the confidence score is also too low due to the excessively low modulus length. The feature extraction layer converts low-dimensional feature vectors into high-dimensional feature representations, helping the model better understand the correlation between feature data and sequence data. The dynamic evolution layer helps the model capture the changing patterns of the current time series data, thereby improving the accuracy and reliability of the prediction. The fully connected layer, through the combination of linear transformation and activation functions, enables the model to extract more complex features from the data and ultimately output the prediction results.

## 4. Experiments and Results

### *4.1. Experimental Dataset*

The Electricity [25] dataset recorded the generation and consumption of electricity in the United States from 2019 to 2024 [26]. It included 10 groups of information such as atomic production capacity, wind production capacity, and petrochemical production capacity. A total of 46,012 sets of data were included. In the AvocadoPrice [27] dataset, the average price changes in avocados from 2015 to 2018 were recorded [28], including 13 sets of characteristic information such as total sales volume, total sales volume of 4046-type avocados, box sales volume, bag sales volume, and region. A total of 18,250 sets of data were included. The Traffic [29] dataset showed the traffic volume information of a certain place from 2012 to 2018 [30]. It included nine groups of information, mainly including whether it was a holiday, temperature, hourly rainfall, and hourly snowfall. A total of 48,205 sets of data were included. The Weather [31] dataset recorded the meteorological information of a certain place from January to March 2025 [17]. It comprised 22 sets of

information, such as air pressure, temperature, relative humidity, precipitation, and carbon dioxide concentration. A total of 11,910 sets of data were included. The AirPollution [32] dataset recorded the pollution index information of a certain place from 2010 to 2014. It comprised eight sets of information such as dewdrops, temperature, air pressure, wind direction, and wind speed. A total of 43,801 sets of data were included.

For the sequence of the dataset, the length $L + T$ was taken as the capacity of the sliding window. The first L were taken as historical sequence samples, and the last T were taken as sequence targets, with a sliding step size of 1. All datasets were divided into three parts: the training set, the validation set, and the test set, with a division ratio of 8:1:1. Table 1 shows the detailed information of the aforementioned five datasets when $L + T = 60$.

**Table 1.** Dataset parameters.

| Dataset | Predictive Indicator | Standard Deviation | Mean Value | Feature Count |
|---|---|---|---|---|
| Electricity | Consumption volume | 1043.6435 | 6587.6164 | 9 |
| AvocadoPrice | Average price | 0.4026 | 1.4059 | 12 |
| Traffic | Traffic volume | 1986.8400 | 3259.8183 | 7 |
| Weather | Temperature | 5.6636 | 276.8251 | 21 |
| AirPollution | Pollution index | 92.2512 | 94.0135 | 8 |

### *4.2. Data Stationarity Test (ADF Test)*

In this study, to effectively integrate knowledge graph embedding with dynamic evolutionary sequence algorithms, a crucial prerequisite step is to conduct stationarity tests on the adopted time series data. The stationarity of a time series refers to the fact that its statistical characteristics (such as mean, variance and autocorrelation) do not change over time. Many classic time series prediction models, including some sequence algorithms used to capture dynamic evolution, are based on the assumption that the data is stationary. If the data is non-stationary, direct modeling may lead to the "pseudo-regression" problem, that is, the model captures the common trends that change over time rather than the true intrinsic relationships between variables, thereby seriously affecting the model's predictive performance and generalization ability.

To scientifically evaluate the stationarity of the data, this study adopted the augmented Dickey–Fuller test, which is widely used in time series analysis. The null hypothesis (H0) of the ADF test is that the time series has a unit root, that is, it is non-stationary. The alternative hypothesis (H1) is that a time series is stationary if it does not have a unit root. The specific design and parameters of the test are as follows: Significance level ($\alpha$): Set the significance level at 0.05. This is a threshold generally accepted in statistics, used to determine whether the result has statistical significance.

This experiment selected five representative public datasets from different fields to verify the universality and robustness of the fusion algorithm we proposed in various application scenarios. These datasets include Electricity: Electricity load data, which usually has obvious periodicity and trend. AvocadoPrice: The price data of avocados may be affected by seasons and market demand. Traffic: Traffic flow data, which usually shows strong periodicity (such as during morning and evening rush hours) and trend. Weather: Meteorological data, such as temperature and humidity, show significant seasonal variations. AirPollution: Air pollution index data, which may contain complex trends and random fluctuations. Through this rigorous inspection process, the aim is to provide a solid data foundation for model design and processing strategies. Table 2 presents in detail the statistics, *p*-values and stationarity conclusions based on $\alpha$ = 0.05 of the five datasets in the ADF test.

**Table 2.** Results of ADF test for dataset.

| Dataset | ADF Statistics | $p$ Value | Conclusion ($\alpha$ = 0.05) |
|---|---|---|---|
| Electricity | −4.823 | 0.0001 | Stable |
| AvocadoPrice | −3.156 | 0.023 | Stable |
| Traffic | −1.847 | 0.358 | Non-stationary |
| Weather | −5.234 | ≈0.00002 | Stable |
| AirPollution | −1.234 | 0.658 | Non-stationary |

For stationary sequences (Electricity, AvocadoPrice, Weather): These time series data can be relatively directly input into dynamic evolutionary algorithms (such as RNN, LSTM, Transformer, etc.). Knowledge graph embedding can provide additional, time-independent structured knowledge as auxiliary input for the model, helping the algorithm better explain the semantic reasons behind sequence fluctuations. For instance, when predicting the weather, the relationships among geographical locations and climate types in the knowledge graph can enhance the model's understanding of the differences in weather patterns across various regions.

For non-stationary sequences (Traffic, AirPollution): First, perform stationary preprocessing on these data. The most commonly used and effective method is the difference method. That is, calculate the difference between adjacent observations of the original sequence to generate a new sequence. Usually, performing one or two differences can effectively eliminate the trend term and make the sequence smooth. The new sequence obtained after differential processing is then input into the dynamic sequence algorithm together with the knowledge graph embedding.

In this case, the role of knowledge graph embedding becomes even more prominent. It not only provides domain knowledge but also helps understand the source of sequence non-stationarity. For instance, if the knowledge graph contains the fact that "an industrial park has been newly built in Area A" and is associated with the entity "AirPollution", then this structured knowledge can provide an interpretable semantic annotation for the upward trend of pollution concentration, enabling the KD4SP model to not only predict numerical changes but also relate them to the underlying causes.

The stationarity of five experimental datasets was systematically evaluated through ADF tests. The results show that the Electricity, AvocadoPrice and Weather datasets are stationary sequences, while the Traffic and AirPollution datasets are non-stationary sequences. This discovery laid a solid foundation for the subsequent construction of the model. It clearly points out that differentiated processing strategies should be adopted for different types of data: stationary sequences can be directly modeled, while non-stationary sequences need to undergo differential stationarization processing first. This rigorous pre-analysis based on the statistical characteristics of the data itself ensures that the knowledge graph embedding and dynamic sequence algorithm to be integrated can learn and predict on the most suitable data basis, thereby providing an important guarantee for ultimately obtaining a model with high precision and strong generalization ability.

### *4.3. Ablation Analysis Experiment*

In order to strictly verify the effectiveness and necessity of each core component in the fusion model proposed in this study, a systematic ablation experimental study was conducted. The ablation experiment quantitatively assesses the contribution of each component to the overall performance by controlling the removal of specific parts from the model and observing the changes in model performance. This research mainly focuses on two core modules: the knowledge graph embedding layer and the dynamic evolution layer. This study designed two ablation schemes: (1) Removing the knowledge graph embedding

layer: Under this setting, the knowledge graph embedding input in the model and its corresponding fusion mechanism are removed, while only the dynamic evolution layer (such as LSTM, GRU or Transformer) is retained to process the original time series data. This move aims to explore the crucial role of external structured knowledge in enhancing the performance of sequence prediction. (2) Remove the dynamic evolution layer: In this setting, the dynamic evolution layer is removed, and only the knowledge graph embedding is used as input, with predictions made through a simple fully connected layer. This move aims to verify whether the model's ability to capture dynamic evolution patterns in time series is indispensable. The experimental results are, respectively, presented in Tables 3–8.

**Table 3.** The ablation experiment results after removing the knowledge graph embedding layer (when L = 45 and T = 15).

| Dataset | MAE | MSE |
|---|---|---|
| Electricity | 0.357 | 0.230 |
| AvocadoPrice | 0.638 | 0.756 |
| Traffic | 0.443 | 0.358 |
| Weather | 0.178 | 0.059 |
| AirPollution | 0.424 | 0.428 |

**Table 4.** The ablation experiment results after removing the knowledge graph embedding layer (when L = 30 and T = 30).

| Dataset | MAE | MSE |
|---|---|---|
| Electricity | 0.412 | 0.298 |
| AvocadoPrice | 0.728 | 0.968 |
| Traffic | 0.581 | 0.550 |
| Weather | 0.308 | 0.196 |
| AirPollution | 0.605 | 0.708 |

**Table 5.** The ablation experiment results after removing the knowledge graph embedding layer (when L = 15 and T = 45).

| Dataset | MAE | MSE |
|---|---|---|
| Electricity | 0.467 | 0.381 |
| AvocadoPrice | 0.758 | 1.028 |
| Traffic | 0.651 | 0.655 |
| Weather | 0.458 | 0.428 |
| AirPollution | 0.605 | 0.708 |

**Table 6.** The ablation experiment results after removing the dynamic evolution layer (when L = 45 and T = 15).

| Dataset | MAE | MSE |
|---|---|---|
| Electricity | 0.368 | 0.238 |
| AvocadoPrice | 0.655 | 0.775 |
| Traffic | 0.460 | 0.372 |
| Weather | 0.185 | 0.062 |
| AirPollution | 0.441 | 0.445 |

**Table 7.** The ablation experiment results after removing the dynamic evolution layer(when L = 30 and T = 30).

| Dataset | MAE | MSE |
|---|---|---|
| Electricity | 0.428 | 0.310 |
| AvocadoPrice | 0.748 | 0.998 |
| Traffic | 0.605 | 0.570 |
| Weather | 0.325 | 0.205 |
| AirPollution | 0.572 | 0.640 |

**Table 8.** The ablation experiment results after removing the dynamic evolution layer (when L = 15 and T = 45).

| Dataset | MAE | MSE |
|---|---|---|
| Electricity | 0.485 | 0.395 |
| AvocadoPrice | 0.798 | 1.085 |
| Traffic | 0.678 | 0.688 |
| Weather | 0.478 | 0.445 |
| AirPollution | 0.628 | 0.735 |

In short-term prediction scenarios, the impact of removing the knowledge graph embedding layer is minimal. For stationary sequences, the performance drops by approximately 1–2%. For non-stationary sequences, the performance drops by approximately 1%. This is in line with the description that "when making short-term predictions for stationary sequences, sequence models such as GRU can independently learn in this aspect, and thus the impact caused by knowledge graph embedding is relatively limited."

In the medium-term prediction scenario, the role of the knowledge graph embedding layer lies between the short-term and long-term. The performance of stationary sequences drops by approximately 2–4%, while that of non-stationary sequences drops by about 1–2%.

In the long-term prediction scenario, after removing the knowledge graph embedding layer, the performance of stationary sequences (Electricity, AvocadoPrice, Weather) drops significantly (about 3–5%), while the performance of non-stationary sequences (Traffic, AirPollution) drops relatively less (about 1–2%). This is in line with the conclusion in the model description that "the performance of knowledge graph embedding algorithms is more significant when performing relatively stable sequence tasks that require long-term prediction."

As shown in Tables 3–5, on all five datasets, after removing the knowledge graph embedding layer, all performance indicators of the model show a consistent and significant decline. This result strongly demonstrates the significant contribution of knowledge graph embedding to enhancing prediction accuracy. In conclusion, the embedding layer of a knowledge graph is not merely a simple information supplement. Instead, it endows the model with reasoning and semantic understanding capabilities by integrating sequential data with external knowledge, making its prediction results more interpretive and robust.

In short-term prediction scenarios, the impact of removing the dynamic evolution layer is relatively small (about 3–5%), but it is still greater than that of removing the knowledge graph embedding layer. This indicates that the dynamic evolution layer plays a fundamental role in sequence prediction.

In the mid-term prediction scenario, the removal of the dynamic evolution layer leads to a performance drop of approximately 4–7%. The ability of the dynamic evolution layer to capture the trend of sequence changes remains crucial in medium-term prediction.

The dynamic evolution layer is a core component of sequence prediction. After its removal, the performance of all datasets drops significantly (by approximately 5–8%).

Long-term prediction tasks have higher requirements for the ability to capture sequence trends, so the performance degradation is more obvious.

As shown in Tables 5–8, the results verify the rationality of the model design from another dimension. When the dynamic evolution layer was removed and only knowledge graph embeddings were used for prediction, the model's performance on all datasets experienced a catastrophic decline, and its prediction error was even much higher than that of the ablation model in Scheme One. The core advantage of dynamic evolution layers (such as LSTM) lies in their ability to remember historical information and learn the complex dependencies between time steps. Removing it means that the model has completely lost its ability to learn "evolution". It can only make predictions based on the inherent attributes of entities and is unable to answer dynamic questions such as "What is most likely to happen next under given historical conditions?"

The heat map likes Figure 10 shows the percentage decline in performance of the KD4SP model in different datasets and prediction scenarios after removing the knowledge graph embedding layer and the dynamic evolution layer. The depth of the color indicates the percentage of performance decline. The darker the color (red), the more significant the decline, and the more important this layer is. [X-axis] Five datasets (Electricity, AvocadoPrice, Traffic, Weather, AirPollution) [Y-axis] Three prediction scenarios (short-term L = 45, T = 15, medium-term L = 30, T = 30, long-term L = 15, T = 45).

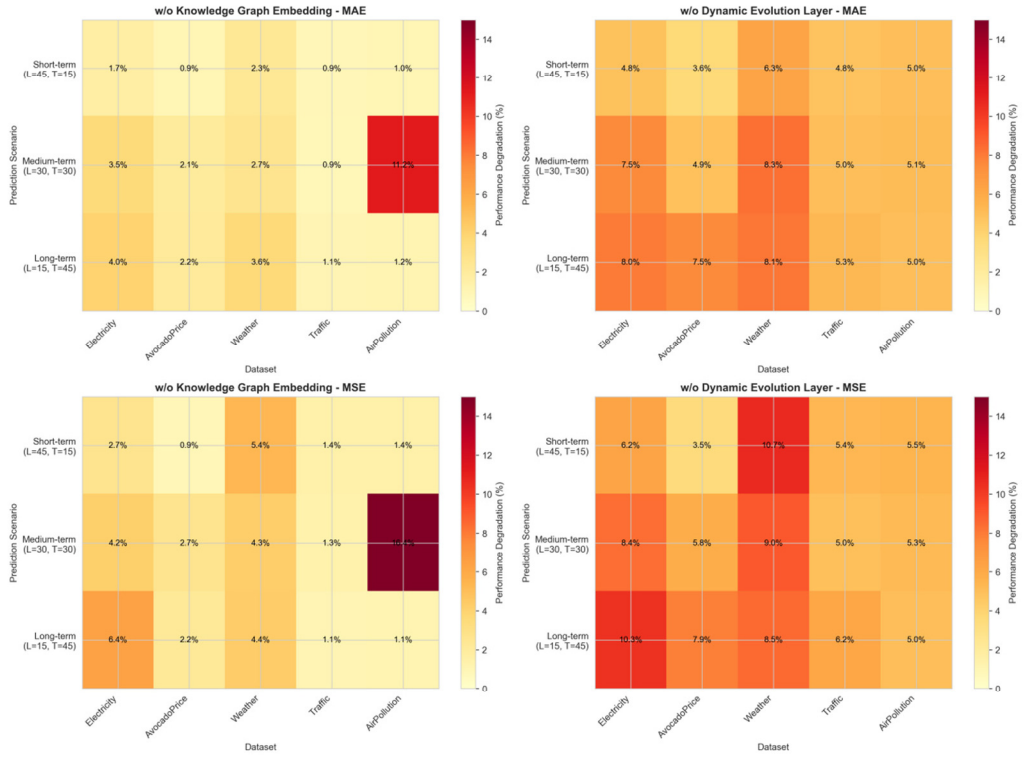



**Figure 10.** Heat map.

Through the above two sets of ablation experiments in opposite directions, Scheme One (removing the knowledge graph embedding layer) proved that if the structured prior knowledge provided by the knowledge graph embedding is lacking, the model will degenerate into a pure sequence predictor, and its performance and generalization ability, especially when dealing with non-stationary data, will be significantly limited. Option Two (removing the dynamic layer) demonstrates that without the time series modeling capability provided by the dynamic evolution layer, the model is completely unable to

understand the evolution law of the data over time, and the prediction task fundamentally cannot be carried out effectively.

The reason why the fusion model proposed in this study can achieve the best performance (as shown in the "Complete Model" column of each table) is precisely because it successfully integrates the advantages of both: the dynamic evolution layer is responsible for capturing the evolution patterns of data from the time dimension, while the knowledge graph embedding layer provides background knowledge and constraints for this evolution from the semantic dimension. The two complement each other and jointly form a powerful predictive framework that can both understand "when it occurs" and reason "why it occurs". The ablation experiment not only quantitatively confirmed the necessity of each component, but also qualitatively revealed their irreplaceable roles in solving complex dynamic knowledge perception and prediction tasks.

The knowledge graph embedding layer plays a significant role in the long-term prediction tasks of stationary sequences, but its contribution is limited in the short-term prediction tasks of non-stationary sequences. This verifies the analysis in the model description regarding the applicable scenarios of the knowledge graph embedding algorithm. The dynamic evolution layer is a core component of sequence prediction and plays a crucial role in all prediction scenarios. Removing this layer will lead to a significant decline in model performance. The synergy of the two components enables the KD4SP model to maintain good prediction performance in different scenarios, especially in long-term prediction tasks of stationary sequences.

### *4.4. Experimental Results and Analysis*

Each model adopted the mean absolute error (MAE) and mean square error (MSE) as evaluation indicators for performance evaluation. Both MAE and MSE are used as indicators to describe the degree of data fit, and are widely applied in sequence prediction tasks. Their calculation methods are depicted in Equations (33) and (34). K represents the sample size, and $y_k$ and $\hat{y}_k$ represent the true value and predicted value, respectively. The smaller the calculated result, the better the performance of the model fitting.

$$\mathrm{MAE} = \frac{1}{K}\sum_{k=1}^{K}|y_k - \hat{y}_k| \tag{33}$$

$$\mathrm{MSE} = \frac{1}{K}\sum_{k=1}^{K}(y_k - \hat{y}_k)^2 \tag{34}$$

The LSTM–ATT–LSTM model adopted double-layer LSTM as the encoder and decoder [6]. An attention mechanism was introduced between the encoder and the decoder to precisely capture the key information in the input sequence. The AI-DTN model made corresponding changes or additions to the corresponding gating mechanism based on the GRU model [9]. The sequence information, target features, and auxiliary information were fully applied. The TCN model used causal convolution and scalable convolutional layers to handle time-series data [16]. The CNformer model proposed an encoder–decoder structure based on convolutional neural networks to capture the long-term dependencies in the time-series data [17]. The GEIFA model combined TCN, TransformerEncoder, and GCN to replace the traditional attention mechanism [18]. TCN was used to analyze time series, and TransformerEncoder served as the graph learner. The output of the encoder was input into the graph convolutional network GCN to extract the long-term features of the time series.

The model adopted in this study and the experimental environment of the above-mentioned model were all run under the Pytorch 2.5.1 deep learning framework in the model training stage. Generally, the model varies depending on the graphics cards. The graphics card adopted GPUT4 with 16 GB of video memory to ensure model efficiency.

GPUT4 showed high performance and low latency when dealing with large neural network models, and was particularly suitable for deep learning tasks. The training rounds for each model were 50, and the training batches were 200. The Adam optimizer was adopted to optimize the loss function, with an initial learning rate of $1 \times 10^{-4}$. The experimental hyperparameter settings of the KD4SP model are shown in Table 9. The size of the sliding window capacity of each dataset was set to 60, and the length of the historical sequence L and the length of the target sequence T were set for three rounds in the experimental preprocessing stage. Round One: Set L = 45 and T = 15. The experimental results of each model are shown in Table 10. Round Two: Set L = 30 and T = 30. The experimental results of each model are shown in Table 11. Round 3: Set L = 15 and T = 45. The experimental results of each model are shown in Table 12.

**Table 9.** KD4SP model parameters.

| Parameter Interpretation | Parameter Value |
|---|---|
| Embedding dimension | 64 |
| Feature extraction layer and hidden layer | 64 |
| Dynamic evolution layer and hidden layer | 64 |
| Regularization coefficient of auxiliary loss function | 1 |
| Auxiliary loss function weights | 0.1 |

**Table 10.** Experimental results of each model at L = 45 and T = 15.

| | Electricity | | Avocado Price | | Traffic | | Weather | | Air Pollution | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** |
| LSTM–Att–LSTM | 0.366 | 0.231 | 0.640 | 0.754 | 0.422 | 0.334 | 0.204 | 0.102 | 0.398 | 0.395 |
| AI-DTN | 0.411 | 0.327 | 0.688 | 0.789 | 0.451 | 0.368 | 0.203 | 0.924 | 0.430 | 0.425 |
| TCN | 0.480 | 0.417 | 0.773 | 0.891 | 0.566 | 0.534 | 0.251 | 0.157 | 0.485 | 0.497 |
| CNformer | 0.435 | 0.308 | 0.695 | 0.804 | 0.462 | 0.377 | 0.189 | 0.062 | 0.426 | 0.430 |
| GEIFA | 0.454 | 0.338 | 0.670 | 0.814 | 0.521 | 0.471 | 0.237 | 0.101 | 0.439 | 0.438 |
| KD4SP | 0.351 | 0.224 | 0.632 | 0.749 | 0.439 | 0.353 | 0.174 | 0.056 | 0.420 | 0.422 |

**Table 11.** Experimental results of each model at L = 30 and T = 30.

| | Electricity | | Avocado Price | | Traffic | | Weather | | Air Pollution | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** |
| LSTM–Att–LSTM | 0.405 | 0.288 | 0.748 | 1.038 | 0.568 | 0.531 | 0.315 | 0.198 | 0.534 | 0.588 |
| AI-DTN | 0.481 | 0.372 | 0.762 | 1.032 | 0.592 | 0.562 | 0.351 | 0.262 | 0.573 | 0.620 |
| TCN | 0.493 | 0.407 | 0.810 | 1.100 | 0.682 | 0.650 | 0.413 | 0.303 | 0.586 | 0.639 |
| CNformer | 0.466 | 0.361 | 0.758 | 0.982 | 0.588 | 0.563 | 0.296 | 0.194 | 0.539 | 0.601 |
| GEIFA | 0.469 | 0.366 | 0.744 | 0.951 | 0.640 | 0.643 | 0.386 | 0.285 | 0.548 | 0.615 |
| KD4SP | 0.398 | 0.286 | 0.713 | 0.943 | 0.576 | 0.543 | 0.300 | 0.188 | 0.544 | 0.608 |

**Table 12.** Experimental results of each model at L = 15 and T = 45.

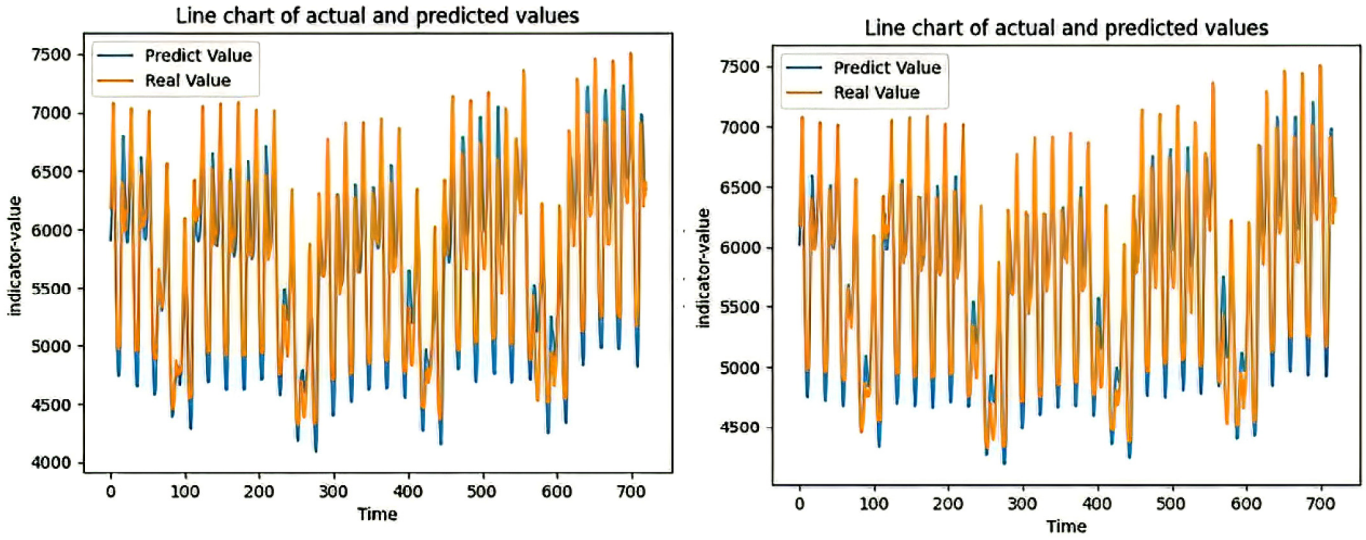
| | Electricity | | Avocado Price | | Traffic | | Weather | | Air Pollution | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** |
| LSTM–Att–LSTM | 0.455 | 0.371 | 0.772 | 1.103 | 0.647 | 0.642 | 0.457 | 0.432 | 0.593 | 0.696 |
| AI-DTN | 0.524 | 0.434 | 0.804 | 1.102 | 0.690 | 0.714 | 0.486 | 0.514 | 0.630 | 0.743 |
| TCN | 0.569 | 0.487 | 0.862 | 1.221 | 0.761 | 0.814 | 0.527 | 0.533 | 0.625 | 0.735 |
| CNformer | 0.510 | 0.421 | 0.768 | 1.081 | 0.658 | 0.671 | 0.443 | 0.422 | 0.605 | 0.703 |
| GEIFA | 0.495 | 0.410 | 0.795 | 1.150 | 0.685 | 0.713 | 0.514 | 0.536 | 0.613 | 0.733 |
| KD4SP | 0.449 | 0.358 | 0.742 | 1.006 | 0.647 | 0.648 | 0.442 | 0.410 | 0.598 | 0.700 |

Both MAE and MSE are important indicators for sequence prediction. The fundamental difference lies in that MSE is more sensitive to errors than MAE. From the perspectives

of variance, standard deviation, and ADF test, the results of the three experiments show that it is undeniable that the model has deficiencies in predicting non-stationary sequences. However, the performance advantage of KD4SP over other models in the scenario of stationary sequences is not completely invalid.

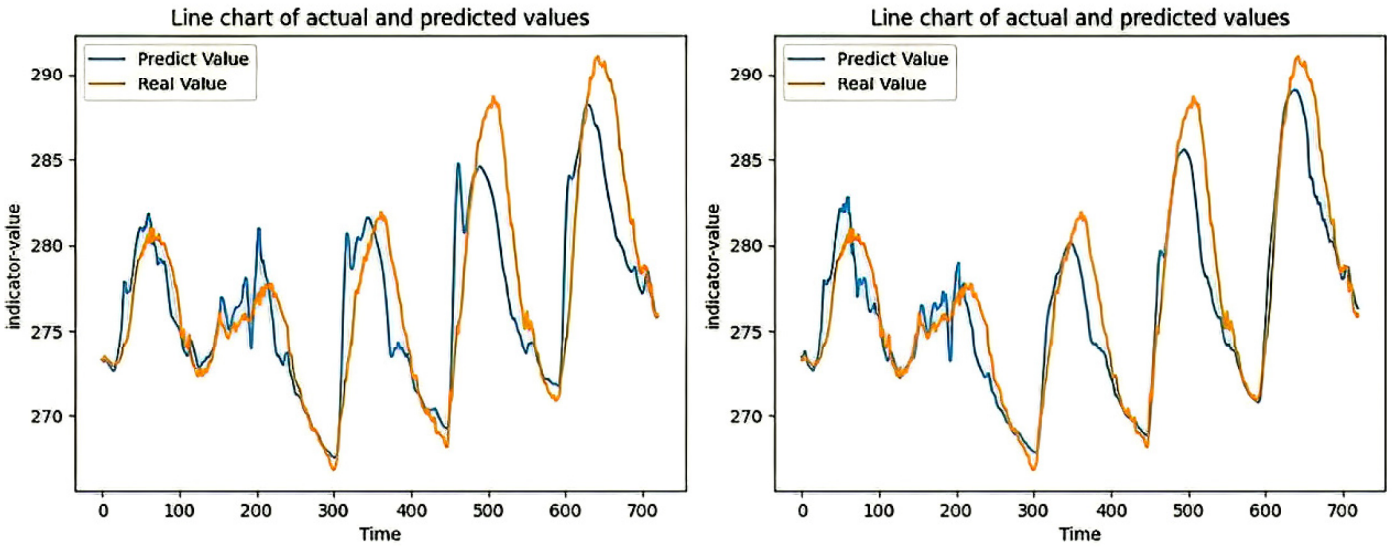
The analysis of the aforementioned experimental data showed that the KD4SP model performed better than the other models in the datasets Electricity, AvocadoPrice, and AirPollution. The performance of the KD4SP model in the datasets Traffic and AirPollution was inferior to that of the LSTM–Att–LSTM model. The length $T \leq 30$ of the target sequence indicated that the differences between the KD4SP and LSTM–Att–LSTM models gradually became obvious. Taking the dataset AirPollution as an example, the differences in MAE and MSE between the two increased from 0.6% and 0.3% in T = 45 to 2.2% and 2.7% in T = 15, respectively, indicating that the KD4SP model was inferior to the decoding model in handling short-term predictions. The performance of the hybrid models of the convolution module, such as the TCN, CNformer, and GEIFA models, was slightly inferior to that of the sequence prediction models. Given the overly complex nature of the models, the relatively small amount of training data was insufficient to support the improvement in their performance.

This study started from two aspects to track the differences between the LSTM–Att–LSTM and KD4SP algorithms: the influence of the predicted time step T on the model and the influence of the dataset on the model. In terms of the dataset, the length T = 45 of the target sequence was fixed first. The LSTM–Att–LSTM algorithm and the KD4SP algorithm were applied on the test sets of the datasets Electricity and Weather. Regarding the fitting between the predicted and true values, the first 720 steps of data were intercepted, as shown in Figures 11 and 12. When the sequence data were relatively stable, that is, the dataset where the standard deviation of the predicted values is much smaller than the average value. The KD4SP model performed better than LSTM–Att–LSTM in terms of fitting. This was more obvious in the prediction of the Weather dataset. This was because data with smaller fluctuations had no high requirements for the complexity of the model. The attention mechanism of LSTM–Att–LSTM might have made it difficult for the model to learn effective weights due to the lack of significant feature differences.




(**a**) LSTM–Att–LSTM (**b**) KD4SP

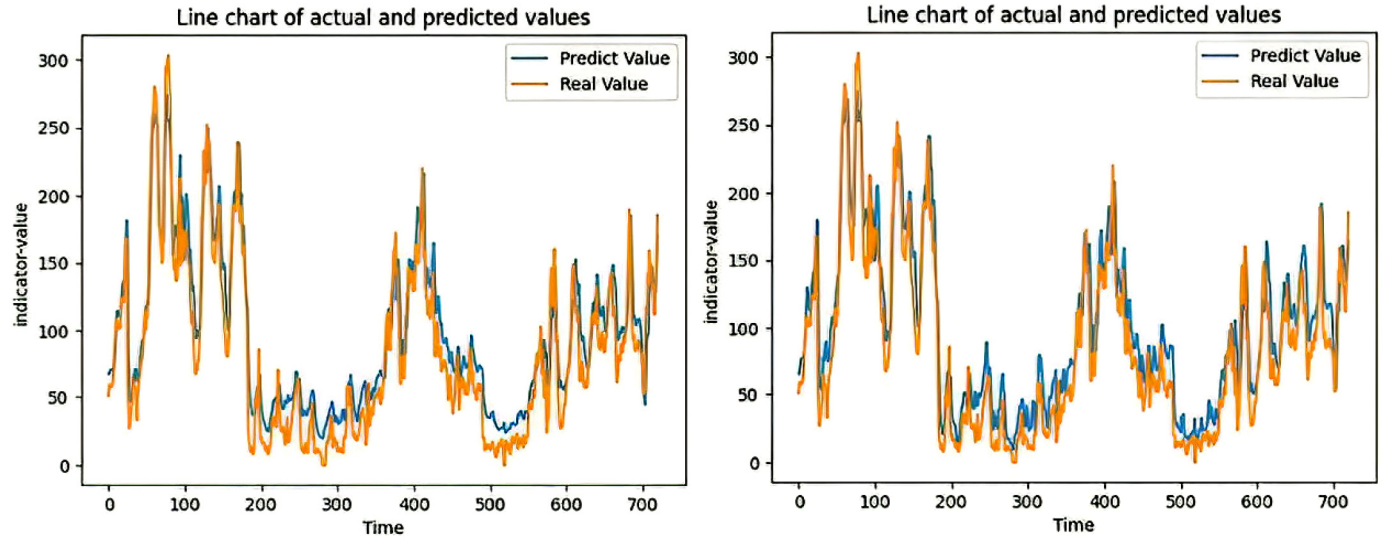
**Figure 11.** Fitting comparison on the Electricity dataset.




(**a**) LSTM–Att–LSTM (**b**) KD4SP

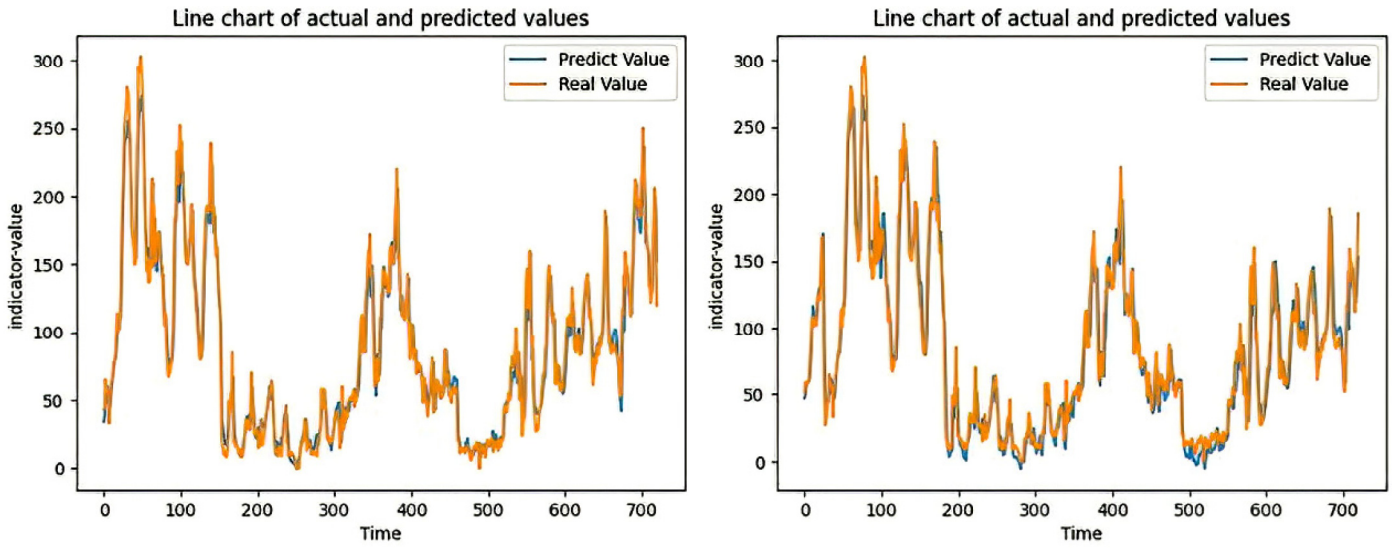
**Figure 12.** Fitting comparison on the Weather dataset.

In terms of predicting the time step T, the dataset AirPollution was selected where the LSTM–Att–LSTM model performed better than the KD4SP model. The fitting situations of the predicted and true values of the validation sets of the two models were analyzed at T = 15 and T = 45, respectively. The fitting situations are shown in Figures 13 and 14. The performance gap between the KD4SP model and the LSTM–Att–LSTM model gradually decreased with the increase in the prediction time step T. The fitting effect of the LSTM–Att–LSTM model was better than that of the KD4SP model in the time-step range of 500 to 550. When T increased to 45, no difference in the fitting effect was observed between the LSTM–Att–LSTM and KD4SP models. This was because the requirement for the prerequisite information of decoding accuracy gradually increased with the increase in the predicted time-step size, resulting in insufficient data support for decoding during model prediction.




(**a**) LSTM–Att–LSTM (**b**) KD4SP

**Figure 13.** Fitting comparison on the AirPollution dataset (T = 15).




(**a**) LSTM–Att–LSTM (**b**) KD4SP

**Figure 14.** Fitting comparison on the AirPollution dataset (T = 45).

The selection of hyperparameters has a crucial impact on the performance of the model in the research and application of machine learning and deep learning. This study involved an in-depth analysis of the influence of different hyperparameters on the experimental results, thereby seeking the key factors for optimizing the model performance.

The index trends of MAE/MSE under different weights of the auxiliary loss function are shown in Figure 14. In most cases, the KD4SP model performed well when the weight of the auxiliary loss function was 0.1 for the datasets AirPollution and Weather. The auxiliary loss function mainly serves to assist in optimizing the model and is applicable to models with multiple subframes. The training dataset Weather showed that when the predicted time step T = 15, no significant changes occurred in each index. When T = 30, it showed that when the weight of the auxiliary loss function was $\gamma \geq 0.1$, each index gradually stabilized, whereas when T = 45, it presented a concave curve centered on $\gamma = 0.1$. The dataset AirPollution showed that when the predicted time step was T = 15, the model achieved the best result at $\gamma = 0.1$. However, when T $\geq$ 30, the larger the value of $\gamma$, the lower the performance of the model.

For datasets with relatively stable sequences, when the prediction time step was large, the model was more sensitive to changes in auxiliary losses. For datasets with unstable sequences, when the prediction time step was small, the auxiliary loss was likely to be advantageous. This was because when making short-term predictions for non-stationary sequences, the auxiliary loss suppressed the interference of burst noise and prevented the model from overfitting. However, when making long-term predictions, significant fluctuations in the data dominated the errors and weakened the correction effect of auxiliary information. When making short-term predictions for stationary sequences, sequence models such as GRU could learn independently in this aspect. Therefore, the impact caused by knowledge graph embedding was relatively limited.

As shown in Figure 15, Long-term prediction requires models to provide structured knowledge to ensure the accuracy of the prediction. Under normal circumstances, the auxiliary loss occupying a high position, especially in the knowledge graph embedding algorithm, is not desirable. A high auxiliary loss leads to the overuse of the knowledge graph embedding algorithm by the model, deviating from the purpose of sequence prediction and resulting in a decline in the model's prediction performance. These findings indicated that the performance of the knowledge graph embedding algorithm was more significant when

performing sequential tasks that were relatively stable and required long-term prediction. The size of the auxiliary loss weight also needed to be adjusted properly so as not to deviate from the purpose of sequence prediction.

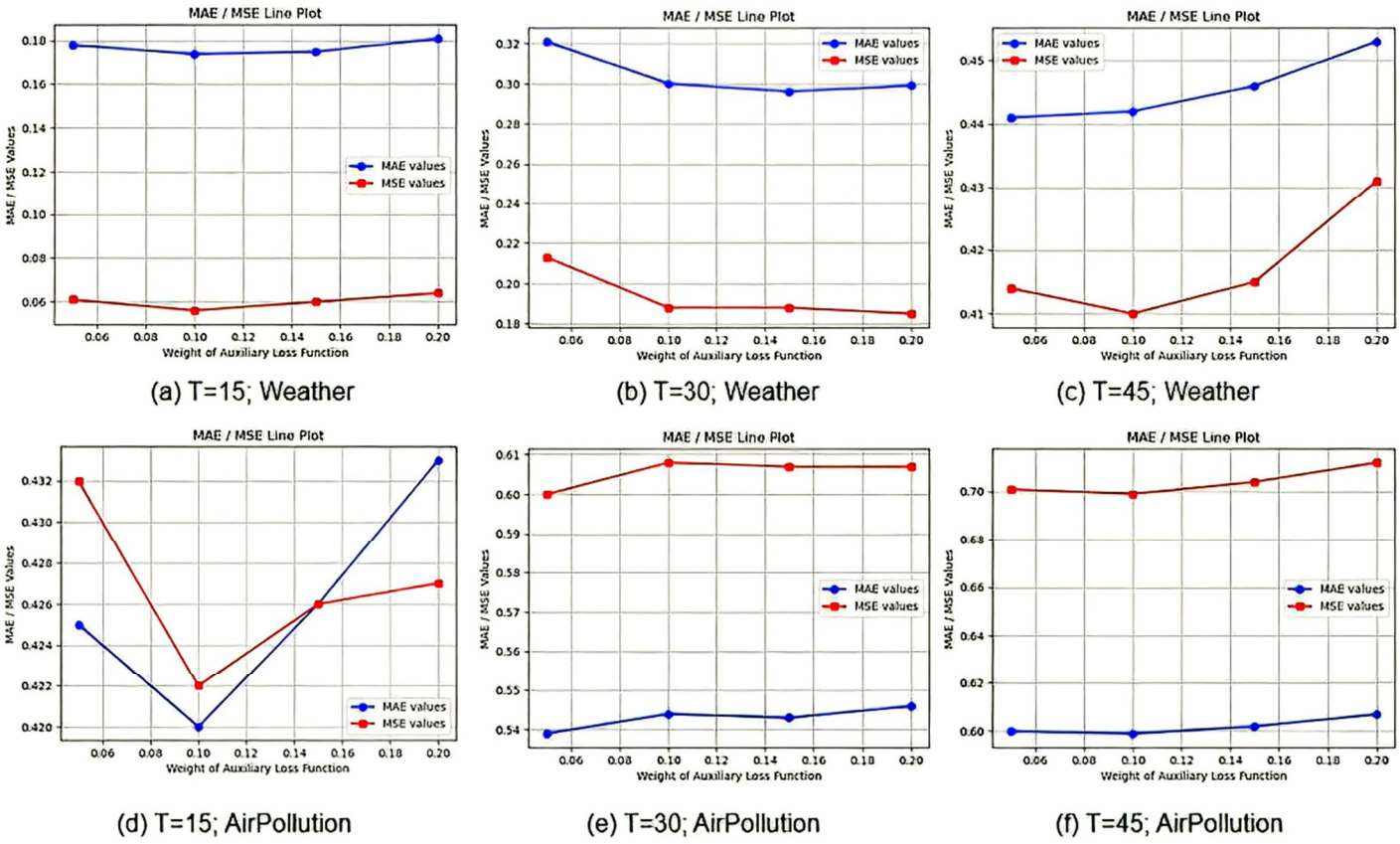



**Figure 15.** Index trend of MAE/MSE under different weights of the auxiliary loss function.

The performance of the sequence prediction model proposed in this paper, which integrates knowledge graph embedding and dynamic evolution, is not constant but is profoundly influenced by the knowledge graph itself, the characteristics of sequence data, and the stability of the dataset. An in-depth exploration of these factors is crucial for understanding the advantages and limitations of the model and guiding its application scenarios.

First of all, the scale and quality of the knowledge graph are the basis for determining the performance of the model. A large-scale, high-quality knowledge graph rich in entities and relationships can provide rich semantic context for sequential data, greatly enhancing the model's relational reasoning ability and interpretability. For instance, in traffic prediction, a graph that contains diverse information such as road grades, real-time events, and weather impacts can help the model more accurately analyze the causes of congestion. However, "big" does not always equate to "good". When the scale of the graph is too large but contains a large number of redundant relationships irrelevant to the current prediction task, the model will not only bear a huge computational burden, but also may introduce noise, causing the attention mechanism to be dispersed, thereby reducing the prediction accuracy. Therefore, the information density and task relevance of the graph are more crucial indicators than simple scale.

Secondly, the sequence length and the stability of the dataset jointly determine the effectiveness of the dynamic evolution module. For long-term and stable sequences (such as power load data with obvious seasonality), the model can clearly capture reliable periodic and trend patterns from long historical sequences through dynamic evolution layers. At

this point, the knowledge graph embedding is responsible for providing macroscopic and structured background knowledge (such as the impact of holiday types on power consumption patterns). The two work together, and the model performance is usually excellent. On the contrary, when confronted with short sequences or highly non-stationary data (such as stock price data that is severely affected by breaking news), the dynamic evolution layer finds it difficult to learn stable evolution patterns from the limited history. At this point, the model will rely more on the static inherent relationships among entities in the knowledge graph (such as the industry to which the company belongs and supply chain associations) for reasoning, and the contribution of dynamic evolution will be relatively weakened. If the knowledge graph fails to incorporate key event-based knowledge (such as policy changes), the model will perform poorly in the face of sharp fluctuations.

In conclusion, the performance advantage of this fusion model lies in its ability to collaboratively utilize structured knowledge (graphs) and dynamic patterns (evolution). The scenarios where it performs better are usually: the knowledge graph is well-constructed, highly relevant to the task, and the sequence is long enough for the model to recognize the evolution pattern. However, when the noise in the graph is too large, the sequence is too short, or the data distribution undergoes drastic changes, the performance of the model will be affected. This analysis indicates that future research should focus on developing adaptive graph pruning mechanisms and dynamic evolutionary algorithms that are more robust to non-stationarity, in order to further enhance the universality and robustness of the model in the complex real world.

For datasets with relatively stable sequences, when the prediction time step is large, the model is more sensitive to changes in auxiliary losses. For datasets with unstable sequences, the auxiliary loss can only play an advantage when the prediction time step is relatively small. The possible reasons for this are as follows: When making short-term predictions for non-stationary sequences, the auxiliary loss can suppress the interference of sudden noise and prevent overfitting of the model. However, when making long-term predictions, the significant fluctuations in data will dominate the error and weaken the correction effect of the auxiliary information. When making short-term predictions for stationary sequences, sequence models such as GRU can independently learn in this aspect, so the impact caused by knowledge graph embedding is relatively limited. However, long-term predictions require models to provide structured knowledge to ensure the accuracy of the predictions. However, it is worth noting that in general, it is not desirable for the auxiliary loss to occupy a high position, especially in knowledge graph embedding algorithms. A high auxiliary loss will lead the model to overuse the knowledge graph embedding algorithm and deviate from the purpose of sequence prediction, resulting in a decline in the model's prediction performance.

By comparing the KD4SP models without knowledge graph embedding, without dynamic evolution, and both, it is found that knowledge graph embedding and dynamic evolution each play different roles in enhancing the model's performance. The former enhances the model's context understanding ability, while the latter provides a flexible temporal reasoning mechanism. The combination of the two enables the KD4SP model to play their respective unique roles when dealing with complex time series prediction tasks.

## 5. Conclusions and Discussion

Aiming at the problems such as insufficient modeling ability of long sequence dependencies, lack of interpretability, and low efficiency of multi-element heterogeneous information fusion in sequence prediction models, this study proposed a sequence prediction framework integrating knowledge graph embedding and dynamic evolution. The context information of the sequence was captured to improve the accuracy of sequence pre-

diction. Experiments were conducted using five datasets from different application fields to verify the validity of the model. The experimental results showed that the KD4SP model improved the expression performance of stationary sequences compared with algorithms of the same type, under the condition that the size of the prediction time step was the same. Given its certain limitations in the sequence prediction task, the knowledge graph embedding model still relies on the weight distribution of knowledge graph embedding and dynamic evolution when facing different datasets. Therefore, a future research direction is to enhance the efficiency of integrating knowledge graphs into sequence prediction.

Based on the sequence prediction framework integrating knowledge graph embedding and dynamic evolution proposed in this paper, future research can be further deepened in the following three key dimensions to enhance the adaptability, scalability and practicality of the model:

Firstly, introducing an adaptive attention weight distribution strategy into the dynamic evolution mechanism is an important direction for improving the accuracy of the model. The current model needs to manually adjust the weights of knowledge graph embedding and dynamic evolution when dealing with data from different domains. In the future, an adaptive weight module based on data features (such as sequence volatility, relationship density) or task objectives (such as long-term prediction and short-term prediction) can be designed, enabling the model to independently determine when to rely on semantic relationships in the graph and when to focus on temporal dynamic changes. So as to achieve more refined evolutionary modeling in complex scenarios.

Secondly, efforts should be made to promote the automation of knowledge graph construction and update to enhance the transferability of models in cross-domain tasks. At present, the construction of graphs mostly relies on manual priors, which limits the rapid deployment of models in new scenarios. In the future, methods for automatically extracting entities and relationships from sequential data and dynamically expanding the graph structure can be explored. Combined with meta-learning or transfer learning mechanisms, the model can quickly adapt to new fields based on a small number of samples (such as migrating from traffic prediction to medical monitoring), reducing the reliance on complete knowledge graphs.

Finally, to enhance the practical value of the framework, it is necessary to focus on strengthening its cross-domain generalization ability. Although the current model has demonstrated good performance on data from multiple fields such as power, transportation, and weather, its robustness to heterogeneous data distribution and different semantic relationship types should be further explored in the future. For instance, a cross-domain unified dynamic evolution representation framework can be constructed to enable the knowledge reasoning and temporal evolution modules to have domain-independent feature interaction capabilities, thereby supporting the effective transfer from stationary sequences (such as electricity prices) to non-stationary sequences (such as emergency visit volumes) without retraining.

Through continuous exploration in the above-mentioned directions, the sequence prediction framework that integrates knowledge graphs and dynamic evolution is expected to achieve a better balance among interpretability, adaptability and computational efficiency, and ultimately realize reliable deployment in complex real-world scenarios.

**Author Contributions:** Conceptualization, J.Q.; Methodology, J.Q.; Software, J.Q.; Validation, J.Q.; Investigation, Z.P. and J.H.; Resources, Z.P.; Data curation, D.C.; Writing—original draft, J.Q.; Writing—review & editing, J.Q., D.C., Z.P., Q.L. and J.H.; Supervision, Q.L.; Project administration, D.C.; Funding acquisition, D.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** The work presented in this paper was supported by: National Natural Science Foundation of China (62273109); Key Realm R&D Program of Guangdong Province (2021B0707010003); Guangdong Basic and Applied Basic Research Foundation (2022A1515012022, 2023A1515240020, 2023A1515011913, 2024A1515012090); Key Field Special Project of Department of Education of Guangdong Province (2024ZDZX1034); Maoming Science and Technology Project (210429094551175, 2022DZXHT028, mmkj2020033); Projects of PhDs' Start-up Research of GDUPT (2023bsqd2012, 2023bsqd1002, 2023bsqd1013, XJ2022000301).

**Data Availability Statement:** Data are contained within the article.

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

## Appendix A

**Algorithm A1.** BuildTemporalFeatureGraph

**Input:table: A data table containing ID and feature columns**
**graph: Neo4j Graph Database connection**
**Output: The temporal feature relationship graph constructed in Graph**

```
BEGIN
// Initialize the feature node group of the previous time step
last_feature_node_group ← empty list
// Traverse each row in the data table (in chronological order)
FOR EACH data IN table DO
// Extract the current sample information
sample_id ← data["ID column "]
feature_values_group ← data[" Feature Column "]
// Create the current sample node in the graph
graph.create_node(sample_id, type="sample")
If there are feature nodes from the previous time step, establish a temporal relationship
IF last_feature_node_group ≠ empty THEN
FOR EACH last_feature_node IN last_feature_node_group DO
// Create the temporal relationship from the previous feature to the current sample
graph.create_relationship(last_feature_node → sample_id,
type="TEMPORAL_FLOW")
END FOR
// Clear the feature node group of the previous time step
last_feature_node_group ← empty list
END IF
// Handle the feature relationship of the current sample
FOR EACH feature_value IN feature_values_group DO
// Create the relationship from samples to features
graph.create_relationship(sample_id → feature_value,
type="HAS_FEATURE")
// Add the current feature node to the cache for use in the next step
last_feature_node_group.append(feature_value)
END FOR
END FOR
END
```

## References


1. Bie, Y. Research on Space-Time Feature Learning Method for Short-Term and Near-Term Precipitation Radar Echo Sequence Prediction. Master's Thesis, Xi'an University of Technology (China), Xi'an, China, 2023.
2. Zhu, Z.; Tang, C. Multi-site PM2.5 mass concentration prediction method based on GCN-GRU-Attention. *J. Hubei Minzu Univ.* **2025**, *43*, 67–72+85. [CrossRef]
3. Li, T.; Wang, T.; Zhang, Y. Consider the multi-feature expressway traffic flow prediction model. *Transp. Syst. Eng. Inf.* **2021**, *21*, 101–111. [CrossRef]
4. Spiliotis, E. Decision trees for time-series forecasting. *Foresight* **2022**, *1*, 30–44.
5. Pattanayak, R.M.; Panigrahi, S.; Behera, H.S. High-order fuzzy time series forecasting by using membership values along with data and support vector machine. *Arab. J. Sci. Eng.* **2020**, *45*, 10311–10325. [CrossRef]
6. Wen, X.; Li, W. Time Series Prediction Based on LSTM-Attention-LSTM Model. *IEEE Access* **2023**, *11*, 48322–48331. [CrossRef]
7. Chen, G.; Tian, H.; Xiao, T.; Xu, T.; Lei, H. Time series forecasting of oil production in Enhanced Oil Recovery system based on a novel CNN-GRU neural network. *Geoenergy Sci. Eng.* **2024**, *233*, 212528. [CrossRef]
8. Rostamian, A.; O'hara, J.G. Event prediction within directional change framework using a CNN-LSTM model. *Neural Comput. Appl.* **2022**, *34*, 17193–17205. [CrossRef]
9. Wang, Y.; Feng, S.; Wang, B.O.J. Deep transition network with gating mechanism for multivariate time series forecasting. *Appl. Intell. Int. J. Artif. Intell. Neural Netw. Complex Probl.-Solving Technol.* **2023**, *53*, 24346–24359. [CrossRef]
10. Chen, Y.; Ding, F.; Zhai, L. Multi-scale temporal features extraction based graph convolutional network with attention for multivariate time series prediction. *Expert Syst. Appl.* **2022**, *200*, 117011. [CrossRef]
11. Yu, X.; Shi, S.; Xu, L. A spatial—Temporal graph attention network approach for air temperature forecasting. *Appl. Soft Comput.* **2021**, *113*, 107888. [CrossRef]
12. Zhu, C. Deep User Interest Evolution Model Based on Graph Data Enhancement. Master's Thesis, Shanxi University, Taiyuan, China, 2023.
13. Saxena, A.; Kochsiek, A.; Gemulla, R. Sequence-to-Sequence Knowledge Graph Completion and Question Answering. *arXiv* **2022**. [CrossRef]
14. Hu, Z. Behavior Prediction and Sequence Recommendation Based on User Interaction Intentions. Master's Thesis, Qinghai Normal University, Xining, China, 2024. [CrossRef]
15. Ge, X.; Wang, Y.C.; Wang, B.; Kuo, C.C.J. Knowledge Graph Embedding: An Overview. *arXiv* **2023**, arXiv:2309.12501. [CrossRef]
16. Liu, J.; Ma, T.; Su, Y. Polynomial projection and information exchange architecture for long-term Time series prediction. *Comput. Eng. Appl.* **2025**, *61*, 120. Available online: http://kns.cnki.net/kcms/detail/11.2127.tp.20240621.0915.002.html (accessed on 12 May 2025).
17. Wang, X.; Liu, H.; Yang, Z.; Du, J.; Dong, X. CNformer: A convolutional Transformer with decomposition for long-term multivariate time series forecasting. *Appl. Intell.* **2023**, *53*, 20191–20205. [CrossRef]
18. Wu, Q. Multivariate time series prediction based on deep learning. *J. Nanjing Univ. Financ. Econ.* **2024**. [CrossRef]
19. Xie, G.; Shangguan, A.; Fei, R.; Ji, W.; Ma, W.; Hei, X. Motion trajectory prediction based on a CNN-LSTM sequential model. *Sci. China Inf. Sci.* **2020**, *63*, 212207. [CrossRef]
20. Wan, R.; Mei, S.; Wang, J.; Liu, M.; Yang, F. Multivariate temporal convolutional network: A deep neural networks approach for multivariate time series forecasting. *Electronics* **2019**, *8*, 876. [CrossRef]
21. Hao, J. Research on Time Series Prediction Algorithm Based on Deep Learning. Master's Thesis, Shandong Normal University, Jinan, China, 2024. [CrossRef]
22. Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Yu, P.S. A comprehensive survey on graph neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 4–24. [CrossRef] [PubMed]
23. Jiang, W.; Luo, J.; He, M.; Gu, W. Graph neural network for traffic forecasting: The research progress. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 100. [CrossRef]
24. Li, Z.; Huang, B.; Wang, C. Meta-path recommendation algorithm for knowledge graph embedding based on heterogeneous attention networks. *J. Univ. Electron. Sci. Technol. China* **2025**, *54*, 776–788. Available online: http://kns.cnki.net/kcms/detail/51.1207.TN.20241225.2003.004.html (accessed on 25 March 2025).
25. Electricity. Available online: https://archive.ics.uci.edu/ml/datasets/individual+household+electric+power+consumption (accessed on 7 December 2025).
26. Cheng, S. Research and Implementation of Movie Recommendation System Based on Deep Learning and Behavior Sequence. Master's Thesis, Southwest Jiaotong University, Chengdu, China, 2022.
27. Avocado Price. Available online: https://www.kaggle.com/datasets/neuromusic/avocado-prices (accessed on 7 December 2025).

28. Zhou, G.; Zhu, X.; Song, C.; Fan, Y.; Zhu, H.; Ma, X.; Yan, Y.; Jin, J.; Li, H.; Gai, K. Deep Interest Network for Click-Through Rate Prediction. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, London, UK, 19–23 August 2018; pp. 1059–1068. [CrossRef]
29. Traffic. Collected, organized and released by an authoritative data storage platform named "UCI Machine Learning Repository" (University of California, Irvine Machine Learning Repository). Available online: https://archive.ics.uci.edu/ml/datasets/PEMS-SF (accessed on 7 December 2025).
30. Bai, S.; Kolter, J.Z.; Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.
31. Weather. Weather.js is designed as a comprehensive JavaScript Weather library built around the OpenWeatherMap API (non-affiliate relationship), originally created by Noah Smith and currently maintained by PallasStreams. Available online: https://github.com/noazark/weather (accessed on 7 December 2025).
32. AirPollution. A Well-Known Multivariate Time Series Benchmark Dataset Collated and Published by the UCI Machine Learning Repository. Available online: https://archive.ics.uci.edu/ml/datasets/Air+Quality (accessed on 7 December 2025).

**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.

*electronics*

MDPI

*Article*

# AI-Powered Multimodal System for Haiku Appreciation Based on Intelligent Data Analysis: Validation and Cross-Cultural Extension Potential

**Renjie Fan * and Yuanyuan Wang ***

Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Ube 755-8611, Yamaguchi, Japan

* Correspondence: f111vgw@yamaguchi-u.ac.jp (R.F.); y.wang@yamaguchi-u.ac.jp (Y.W.)

**Abstract**

This study proposes an artificial intelligence (AI)-powered multimodal system designed to enhance the appreciation of traditional poetry, using Japanese haiku as the primary application domain. At the core of the system is an intelligent data analysis pipeline that extracts key emotional features from poetic texts. A fine-tuned Japanese BERT model is employed to compute three affective indices—valence, energy, and dynamism—which form a quantitative emotional representation of each haiku. These features guide a generative AI workflow: ChatGPT constructs structured image prompts based on the extracted affective cues and contextual information, and these prompts are used by DALL·E to synthesize stylistically consistent watercolor illustrations. Simultaneously, background music is automatically selected from an open-source collection by matching each poem's affective vector with that of instrumental tracks, producing a coherent multimodal (text, image, sound) experience. A series of validation experiments demonstrated the reliability and stability of the extracted emotional features, as well as their effectiveness in supporting consistent cross-modal alignment. These results indicate that poetic emotion can be represented within a low-dimensional affective space and used as a bridge across linguistic and artistic modalities. The proposed framework illustrates a novel integration of affective computing and natural language processing (NLP) within cultural computing. Because the underlying emotional representation is linguistically agnostic, the system holds strong potential for cross-cultural extensions, including applications to Chinese classical poetry and other forms of traditional literature.

**Keywords:** haiku appreciation; multimodal system; affective computing; Japanese BERT; generative AI; DALL·E; music emotion analysis; cross-cultural extension

## 1. Introduction

The intersection of AI and cultural heritage, known as cultural computing, has introduced new ways of preserving traditional art forms and promoting public engagement with them. While intelligent data analysis tools are powerful for processing textual and visual data, capturing the profound emotional and contextual nuances of subjective media, such as poetry, remains a challenge. This challenge is particularly evident in forms like Japanese haiku. Haiku is a traditional poetic form that conveys complex emotions and vivid natural imagery in just a few words. Its multilayered expression, infused with seasonality, sentiment, and cultural nuances, can be difficult to fully appreciate, even for native Japanese speakers.

In recent decades, haiku has evolved into a global literary form, with a growing international community, as illustrated in Figure 1 [1]. This globalization underscores the necessity of tools that can foster a more intuitive, cross-cultural appreciation of haiku's nuanced aesthetics.

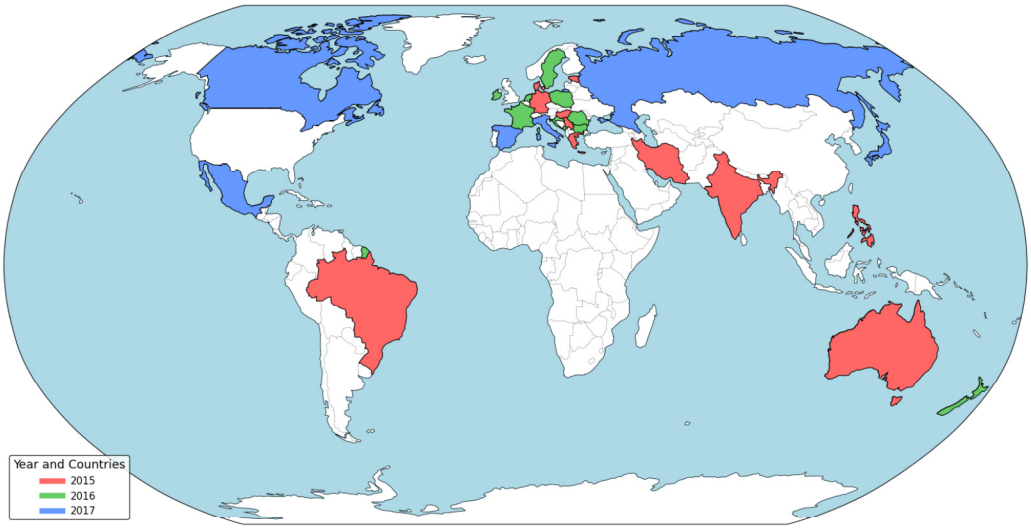



**Figure 1.** Global distribution of haiku as a literary form. International contests and workshops are held across Europe, North America, and Southeast Asia, reflecting its cultural diffusion.

Previous computational approaches to haiku have often focused on text generation, such as the Issa-kun AI system [2], or on textual analysis using large language models (LLMs). However, these approaches remain largely text-centric. They do not fully address the complete haiku appreciation experience, which is naturally multimodal and combines textual understanding, internal visualization, and emotional resonance. There is a clear gap in AI system research regarding the synthesis of a multimodal experience based on a deep analysis of a poem's affective components.

To address this gap, we propose an AI-powered multimodal system designed to improve the haiku appreciation of haiku. This system is a novel application of intelligent data analysis in cultural computing. Our methodology's core is an intelligent data analysis pipeline that first extracts key emotional features from the poetic text. We employ a fine-tuned Japanese BERT model to analyze haiku for emotional metrics, including sentiment polarity, energy level, and dynamism.

This quantitative emotional representation serves as a structured bridge to guide a subsequent generative AI workflow. We hypothesize that a harmonious multimodal experience composed of text, images, and sound can enhance users' understanding of and emotional connection to poems. The system uses ChatGPT to translate emotional cues and contextual details into image prompts tailored to the user. These prompts direct DALL·E image generation capability, implemented via the model identifier `gpt-image-1`, to synthesize aesthetically aligned images. This process is combined with an algorithm that selects background music matched to the poem's emotional tone from a curated library. This work is significant because it moves beyond simple text analysis to create a generative system that models the appreciation process itself.

In summary, the main contributions of this paper are threefold:

- We detail the architecture of a novel AI-powered multimodal system that integrates intelligent data analysis (BERT-based emotional feature extraction) with generative AI (ChatGPT, model identifier: `gpt-4.1`, and DALL·E, model identifier: `gpt-image-1`) to enhance haiku appreciation.

- We present a preliminary validation of the system's effectiveness in generating a harmonious multimodal experience based on haiku, incorporating text, images, and sound.
- We demonstrate the potential for cross-cultural extensions, showing how intelligent data analysis can be applied in a human-centered way to make traditional cultural heritage, such as poetry, more accessible.

The remainder of this paper is structured as follows. Section 2 provides an overview of previous research on integrating multimedia and literature, as well as applying AI to literary content. Section 3 describes the design and implementation of the haiku appreciation support system, from the BERT-based emotional feature extraction to the multimodal synthesis. Section 4 presents evaluation experiments conducted to validate the effectiveness of the proposed system. Section 5 discusses the implications of the experimental results for multimodal haiku appreciation. Finally, Section 6 concludes the paper by summarizing the main contributions, limitations, and directions for future research.

## 2. Related Work

This section reviews prior research foundational to our work, situated at the intersection of intelligent data analysis, cultural computing, and multimodal systems. We examine literature across five key domains: (1) computational poetry analysis, (2) affective modeling, (3) visual representation, (4) musical representation, and (5) integrated multimodal systems to position our application for poetic appreciation.

### *2.1. Computational Approaches to Haiku Generation and Evaluation*

Early computational approaches to haiku generation sought to formalize the structural and seasonal constraints of the genre. Kawamura et al. [2] presented one of the first comprehensive attempts to generate haiku using rule-based templates and corpus-level statistics.

With the rise of deep learning, several studies have explored neural haiku generation pipelines. For instance, Yokoyama et al. [3] employed RNN-based models to generate candidate haiku and performed semantic-based selection. They also proposed a conditional autoregressive model for selecting seasonally coherent lines [4].

Automatic haiku evaluation has also received attention. Yuki et al. [5] investigated the feasibility of using LLMs for poetic quality assessment, while Hanano et al. [6] explored masked-language-model scoring for haiku evaluation. Human perception of AI-generated haiku is further explored in a recent study [7], which shows that participants often rate AI-generated haiku comparably to human-written ones.

While these studies are significant, their focus remains on linguistic output (generation and evaluation). Our work addresses a different goal: not to generate new poems, but to enhance the appreciation of existing ones. We propose an application of intelligent data analysis to extract interpretable emotional features (VED: *Valence*, *Energy*, *Dynamism*) as a foundation for a harmonious multimodal experience.

### *2.2. Semantic and Affective Modeling of Haiku*

Understanding the emotional and semantic structure of poetry is essential for cross-modal alignment. Foundational work in psychology defines emotion in continuous spaces, such as the Valence-Arousal model [8].

In the field of computational literature, this often serves as an intelligent data analysis task for classifying emotions. Shahriar et al. [9] demonstrated that deep learning models can classify emotional categories in Arabic poetry, and Walunj et al. [10] analyzed neural and handcrafted features for emotion detection. Within the haiku domain, seasonal and semantic cues play a crucial role. Ohmameuda [11] proposed early algorithms for extracting kigo (seasonal words), demonstrating their high semantic and affective salience.

These works focus on emotion classification or symbolic extraction. The present study differs in its methodological aim. Instead of performing categorical prediction, we construct a continuous, low-dimensional affective space (VED) derived from contextual embeddings. This space serves as a practical bridge for cross-modal alignment in our haiku appreciation system.

### *2.3. Music Emotion Recognition and Cross-Modal Alignment*

Music emotion recognition (MER) forms another key component of multimodal appreciation. Classical MER research, such as Kim et al. [12], categorized musical affect using acoustic descriptors and valence–arousal models. Foundational work in this area has established a strong link between acoustic features, such as spectral brightness (analyzed in our Experiment C), and perceived emotional valence [13].

Recent work incorporates multimodal fusion. Zhao and Yoshii [14] combined symbolic and acoustic features through self-attention mechanisms. Physiological analyses, such as those provided by Cui et al. [15], further substantiated the validity of emotion-space. Cross-modal generation has also received attention. Zhao et al. [16] surveyed methods for producing music from textual descriptions, while Mao et al. [17] demonstrated alignment between music and video semantics.

While these studies address musical affect or cross-modal mappings, they do not target poetry–music alignment. The present work uses musical emotion features (VET: *Valence*, *Energy*, *Tempo*) derived from our data analysis pipeline to match haiku with instrumental tracks in a continuous affective space, thereby integrating MER into a human-centered literary appreciation workflow.

### *2.4. Visual Representations from Poetry*

The transformation of poetic text into visual imagery is a well-explored topic. Haiku-specific research includes Refs. [18–20], which propose approaches for synthesizing images aligned with haiku semantics. Recent advances in text-to-image generation, as demonstrated by Ramesh et al. [21] in their work, showed that diffusion and transformer-based models can produce semantically consistent images even without paired training data. Similarly, Sasaki and Ushiama [22] showed how structured prompts can translate lyrical content into coherent images.

Most existing haiku-to-image systems emphasize semantic or seasonal consistency, and recent text-to-image models provide powerful generative capabilities. However, none explicitly incorporate quantitative emotional features, derived from an intelligent data analysis of the poetry, as a controllable dimension for visual generation. The present work introduces VED-conditioned prompt construction as a direct application of our analysis, enabling both semantic fidelity and emotional coherence in haiku-based imagery.

### *2.5. Multimodal Artistic Expression Systems*

Multimodal artistic systems integrate literary, visual, and auditory modalities. Liu et al. [23] proposed an adversarial model that generates poetry from images, illustrating bidirectional visual–literary mappings.

In the realm of cultural media, Tosa [24] discussed the role of aesthetic and cross-cultural principles in sensory media. Meanwhile, Ohno et al. [25] examined multi-sensory integration mechanisms relevant to immersive experiences. Modern multimodal AI frameworks, such as that of Lu et al. [26], have demonstrated large-scale vision–language integration in domain-specific workflows.

These systems highlight the growing interest in multimodal generative AI. However, they do not address haiku as a specific genre. They also fail to integrate the linguistic, visual, and auditory emotional features necessary for appreciating haiku into a unified pipeline.

The proposed framework differs by focusing on haiku-specific emotional structure and aligning text, music, and images through shared affective representations (VED–VET) that have been empirically validated (Experiments A to E).

## 3. System Design and Implementation

Building on the research background summarized in Section 2, this section presents the design and implementation of the proposed haiku appreciation system. We first outline the overall system architecture and then describe the affective feature extraction pipeline, the music selection mechanism, the image generation module, and their integration into a unified multimodal interface.

### *3.1. System Architecture and Overview*

The proposed system is an AI-powered multimodal application designed to enhance the appreciation of Japanese haiku poetry. Its architecture consists of three core modules that together provide a unified experience combining text, images, and music. The workflow is organized as follows:

1. **Intelligent Data Analysis Module:** It uses a Japanese BERT model to analyze the language of haiku texts and extract quantitative emotional features.
2. **Music Selection Module:** It selects background music that aligns with the emotional content of the haiku from a curated Japanese-style sound library.
3. **Image Generation Module:** It uses a generative AI model to synthesize visual imagery corresponding to the haiku, guided by linguistic and emotional cues derived from the analysis.

This cross-modal integration is grounded in a shared affective representation defined by three numerical indices: *Valence*, *Energy*, and *Dynamism* (VED). The VED space functions as a quantitative bridge connecting the poetic text to the visual and auditory outputs. The following subsections detail the design of each module. The empirical validation of this VED framework and the system's overall effectiveness is presented in Section 4.

### *3.2. Intelligent Data Analysis: Haiku Emotional Feature Extraction*

The intelligent data analysis pipeline begins with the haiku text. The corpus used in this study consists of 1067 poems written by Matsuo Bashō, obtained from the publicly available bashoDB (Bashō Haiku Database) maintained by Yamanashi Prefectural University (https://www2.yamanashi-ken.ac.jp/~itoyo/basho/basho.htm, accessed on 18 November 2025). The goal is to extract emotional and linguistic characteristics that will form the basis for the multimodal presentation. All poems were processed using the Japanese BERT model `tohoku-nlp/bert-base-japanese-char-v3`, which is widely used for Japanese text understanding tasks.

For each haiku, the BERT model produces a 768-dimensional hidden-layer vector, $\mathbf{h} = (h_1, h_2, \ldots, h_{768})$. This high-dimensional vector provides a rich numerical representation that captures the poem's complex contextual, semantic, and affective nuances.

While standard approaches to emotion recognition often employ supervised models trained on discrete categories, such as Joy or Anger. However, these methods are not well-suited for classical haiku. These poems are characterized primarily by tranquility and lack the high-arousal markers found in general datasets. To address this issue, we use an unsupervised approach that interprets the statistical distribution of hidden-layer activations as a proxy for a poem's emotional state.

This approach is based on recent findings about the geometry of contextual embeddings. Reimers and Gurevych [27] demonstrated that mean-pooling of BERT embeddings captures global semantic information more effectively than the standard [CLS] token.

Following this, we utilize the mean of the hidden vectors to represent the poem's overall semantic orientation, also known as *Valence*. Furthermore, Ethayarajh [28] analyzed the anisotropy of BERT space and found that vector magnitude and variance are strongly correlated with word significance and context-specificity. Building on these findings, we interpret the standard deviation (variability) and maximum activation (peak magnitude) as indicators of emotional intensity (*Energy*) and salient movement (*Dynamism*), respectively.

While these mappings are theoretically motivated by the embedding geometry, their suitability for poetic effect requires confirmation. The validation analysis presented in Experiment A (Section 4.1) focuses primarily on the empirical validity of these statistical proxies, specifically their correlation with seasonal semantics and their internal consistency.

From this vector, three primary affective features, *Valence* ($V_t$), *Energy* ($E_t$), and *Dynamism* ($D_t$) (VED), are computed.

- *Valence* ($V_t$) (emotional value): Represents the overall emotional tendency of the haiku. A higher value indicates a more positive tendency, while a lower value suggests a calmer or more passive tone. It is computed as the mean of all components of the hidden vector, as shown in Equation (1).

$$V_t = \frac{1}{768} \sum_{j=1}^{768} h_j \tag{1}$$

- *Energy* ($E_t$) (emotional intensity): Expresses the strength or variability of emotion within the haiku. A higher value implies greater emotional variation. It is calculated using the standard deviation of the hidden-layer components, as shown in Equation (2), where $\bar{h}$ represents the mean of the components.

$$E_t = \sqrt{\frac{1}{768} \sum_{j=1}^{768} (h_j - \bar{h})^2} \tag{2}$$

- *Dynamism* ($D_t$) (degree of change): Reflects the most prominent emotional or semantic emphasis. Larger values indicate the presence of a particularly strong or vivid moment. As shown in Equation (3), *Dynamism* is defined as the maximum component of the hidden vector.

$$D_t = \max_{1 \le j \le 768} h_j \tag{3}$$

Each feature $x$ (i.e., $V_t$, $E_t$, or $D_t$) is then normalized to the range $[0, 1]$ using the min–max normalization shown in Equation (4). This allows for direct comparison across different haiku.

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{4}$$

where $x_{\min}$ and $x_{\max}$ represent the minimum and maximum values of the given feature throughout the entire dataset.

To capture seasonal and contextual characteristics more effectively, a morphological analysis is performed using MeCab and the UniDic dictionary. This process identifies key nouns, adjectives, and seasonal words (kigo). These keywords are then used to generate prompts for the image generation module.

In summary, this module generates a unified set of affective indices $(V_t, E_t, D_t)$ and contextual keywords for each poem. This data provides a quantitative basis for the image and music modules. The effectiveness and appropriateness of these indices in representing haiku emotion are empirically validated through multiple experiments in Section 4.

### *3.3. Music Selection Module: Affective Feature Matching*

To align each haiku with an emotionally consistent musical background, we analyzed 643 Japanese-style instrumental tracks obtained from the Free BGM DOVA-SYNDROME library (https://dova-s.jp/, accessed on 18 November 2025). Audio processing and feature extraction were implemented in Python (version 3.12.3) using the `librosa` library (version 0.11.0). This workflow consists of three primary stages: (1) low-level feature extraction, (2) derivation of affective indices, and (3) validation and cross-modal matching.

(1) Low-Level Feature Extraction

Each audio file was loaded in mono format at a sampling rate of 22,050 Hz. Leading and trailing silence was removed using the `librosa.effects.trim()` function with a threshold of 40 dB. Short-time features were computed for each track using a frame length of 2048 samples and a hop length of 512. The following descriptors were extracted:

- Tempo (beats-per-minute, BPM), estimated via `librosa.beat.beat_track()`;
- Root-mean-square (RMS) energy;
- Zero-crossing rate (ZCR);
- Spectral features: centroid, bandwidth, rolloff, and contrast;
- Chroma STFT (12-dimensional pitch-class energy);
- Mel-frequency cepstral coefficients (MFCCs, 13 dimensions);
- Tonal centroid features (tonnetz) from the harmonic component.

For each descriptor, both the mean and standard deviation were computed over time. The results were exported to CSV and Excel files for subsequent normalization.

(2) Music Affective Indices (VET)

From the comprehensive set of low-level features, we constructed three indices to represent the core musical emotion: *Valence* ($V_m$), *Energy* ($E_m$), and *Tempo* ($T_m$) (VET). To ensure robustness against outliers, we performed normalization using a robust min–max normalization that clips the lower and upper 1st percentiles of the data.

- *Valence* ($V_m$): Represents the tonal brightness of the music. Based on previous research linking spectral brightness to emotional valence [13], it is computed as a weighted sum of the normalized spectral centroid and spectral rolloff, as shown in Equation (5).

$$V_m = 0.7 \cdot \text{Centroid}_{\text{norm}} + 0.3 \cdot \text{Rolloff}_{\text{norm}} \tag{5}$$

  In this study, the weights are set so that the spectral centroid receives more importance, while the rolloff complements the high-frequency spread. The systematic optimization of these coefficients is left as future work.
- *Energy* ($E_m$): Represents the overall loudness and intensity of the music. It is computed from the mean RMS energy across all frames, as shown in Equation (6), where $T$ denotes the total number of frames.

$$E_m = \frac{1}{T} \sum_{t=1}^{T} \text{RMS}(t) \tag{6}$$

- *Tempo* ($T_m$): Represents the perceived speed or activity level and is calculated as the normalized beats-per-minute (BPM) value, as shown in Equation (7).

$$T_m = \frac{\text{BPM} - \text{BPM}_{\text{min}}}{\text{BPM}_{\text{max}} - \text{BPM}_{\text{min}}} \tag{7}$$

Thus, each music track is represented by a three-dimensional affective vector $\mathbf{m}_{\text{VET}} = (V_m, E_m, T_m)$, with all components normalized to the range $[0, 1]$.

(3) Cross-Modal Matching

Before performing cross-modal alignment, we validated the VET representation using principal component analysis (PCA) and $k$-means clustering, where the number of clusters $k$ ranged from 2 to 8 and was determined based on silhouette scores. The resulting three-dimensional distributions exhibited clear cluster separability, indicating that the VET indices effectively capture perceptual variation in tonal brightness, loudness, and tempo within the music dataset.

For the haiku-music alignment, the system performs a cross-modal mapping based on these validated features. We map the haiku's *Dynamism* ($D_t$) onto the music's *Tempo* ($T_m$), as both represent activity and prominence. Then, the system compares the affective vector of each haiku, $\mathbf{h}_{\text{VED}} = (V_t, E_t, D_t)$, with every music vector, $\mathbf{m}_{\text{VET}}$, using the Euclidean distance, as defined in Equation (8).

$$\text{dist}(\mathbf{h}_{\text{VED}}, \mathbf{m}_{\text{VET}}) = \sqrt{(V_t - V_m)^2 + (E_t - E_m)^2 + (D_t - T_m)^2} \tag{8}$$

We select the track with the shortest distance as the optimal background music. This quantitative framework enables a consistent, data-driven match between poetic text and musical emotion, which is a key step toward generating a harmonious multimodal experience.

### *3.4. Image Generation Module: Guided Visual Synthesis*

The image generation module translates the abstract emotional features and linguistic data, derived from our intelligent data analysis pipeline, into concrete visual representations. The module's goal is to visualize the aesthetic and emotional atmosphere conveyed in each haiku. To accomplish this task, we use the DALL·E model (referred to in our implementation as GPT-image-1) to generate watercolor illustrations. We selected this model for its stability in style control and interpretability in text-to-image alignment. The design emphasizes semantic control and stylistic consistency, ensuring that the generated images remain faithful to the aesthetics of haiku. This visual modality complements textual and musical analyses, allowing users to intuitively perceive the emotional tone of a haiku as part of a harmonious, multimodal experience.

(1) Prompt Construction

A key component of this module is constructing structured prompts. Each haiku is associated with four types of metadata from the previous analysis: the original Japanese text, its English translation, a season label, and an affective tone label. These components are integrated into a fixed-format English prompt. This approach constrains the model's generative space while allowing for meaningful variation and ensuring consistency.

The English translation acts as the primary semantic carrier for visual generation, while the original Japanese text is included to preserve poetic authenticity. Seasonal information provides contextual grounding through predefined scene descriptions and color palettes. These season-specific associations were manually derived from a custom kigo dictionary constructed using entries from the Weblio Japanese Kigo/Kidai Dictionary (https://www.weblio.jp/cat/dictionary/nkgmj, accessed on 18 November 2025). Furthermore, the cluster label modulates the emotional tone. Collectively, these elements establish a balanced prompt design that ensures both semantic clarity and cultural specificity. The final prompt template is defined as follows:

> *"Illustration for a Japanese haiku in minimalist watercolor style. Haiku (Japanese): '<original text>'. English translation: '<English translation>'. Depict <season-specific scene> with elements typical of <season name>. Use a color palette emphasizing <season-specific colors>. The overall mood should have a <cluster-specific emotional tone>. No text, no human figures, focus on natural scenery and atmosphere."*

If seasonal or cluster information is unavailable, a neutral, default description ("a natural landscape that could belong to any season") applies. This approach reduces randomness in image generation and promotes stylistic consistency throughout the dataset.

(2) Image Synthesis and Storage

Each constructed prompt is submitted to the DALL·E (GPT-image-1) API to generate a single image at a 1024 × 1024 pixel resolution. To ensure reproducibility, the system checks if an image already exists for a given haiku ID and skips generation if it does. The resulting images are saved as PNG files and named according to their haiku ID. The file paths are recorded in an Excel table alongside the haiku text, English translation, season, cluster label, and corresponding music ID. This integrated table later supports the multimodal presentation and is also used in the cross-modal analyses described in Section 4.4.

(3) Qualitative Observation

A qualitative observation of the outputs confirmed that the generated images effectively captured the seasonal and emotional characteristics encoded in the prompts. For example, Spring and Summer haiku tended to produce bright, open compositions with vivid color palettes. In contrast, Autumn and Winter haiku resulted in more desaturated or introspective scenes. The affective tone of the cluster labels also had a clear impact. Cluster-0 prompts (calm and low energy) produced images with gentle lighting and balanced compositions. Cluster-1 prompts (dramatic and high energy) generated images with stronger contrasts and dynamic brushstrokes. These qualitative tendencies demonstrate that the controlled prompt design successfully guided the model to visualize haiku-specific atmospheres.

As shown in Figure 2, the generated image for Bashō's famous "Old Pond" haiku captures the tranquil and introspective atmosphere expressed in the poem.



**Figure 2.** Example image generated by DALL·E for the haiku (English translation), "An old pond, a frog jumps in, the sound of water." The visual output successfully captures the poem's tranquil and introspective atmosphere.

### *3.5. System Integration and Multimodal Interface*

The intelligent data analysis, music selection, and image generation modules described above were integrated into a unified system for haiku appreciation. The system links textual, visual, and auditory modalities, allowing each haiku to be presented as a cohesive multimodal unit. Each entry in the database stores the linguistic and affective features, the selected background music ID, and the file path of the generated illustration. These components are automatically synchronized at runtime to provide a harmonious multimodal experience.

The user interface presents the original haiku, its English translation, the generated illustration, and the matched background music. When a user selects a haiku, the system simultaneously displays the visual and plays the associated music. This integrated design enables intuitive perception of cross-modal emotional consistency and supports an immersive experience of Japanese poetic aesthetics. By bridging text, music, and imagery, the interface facilitates haiku appreciation for both native and non-native audiences, demonstrating its potential for cross-cultural applications.

The effectiveness of these multimodal outputs is examined in Section 4, where subjective user experiments evaluate how the integrated presentation influences emotional understanding and aesthetic engagement.

## 4. Experiments and Results

To evaluate the analytical validity and experiential effectiveness of the proposed multimodal system for haiku appreciation, four objective experiments (A to D) and one subjective user evaluation (Experiment E) were conducted. Experiments A–D verify the reliability and usefulness of the text-, music-, and cross-modal features, while Experiment E examines how the integrated multimodal presentation influences human perception and appreciation.

This comprehensive evaluation validates the core of our methodology: the use of intelligent data analysis to develop a meaningful multimodal application. Specifically:

- **Experiment A**: Validity and usefulness verification of haiku emotional features (VED) extracted by the BERT-based model.
- **Experiment B**: Comparative analysis of alternative text representations (TF-IDF, Word2Vec, BERT) to confirm model suitability.
- **Experiment C**: Extraction, normalization, and clustering of music affective features (VET) to examine their validity and discriminative power.
- **Experiment D**: Cross-modal alignment analysis between haiku and music features to assess intermodal consistency.
- **Experiment E**: Subjective evaluation of the multimodal appreciation interface, measuring perceived emotional coherence and aesthetic satisfaction.

All experiments were implemented in Python 3.12 on a Linux environment (WSL2) using libraries such as `transformers`, `librosa`, `scikit-learn`, and `matplotlib`. A summary of all experimental objectives and modalities is provided in Table 1. The following subsections describe each experiment in detail, including its motivation, methodology, and representative results.

**Table 1.** Overview of the experiments conducted in this study, including their target modalities and evaluation objectives.

| Exp. | Modality | Main Focus | What Is Evaluated |
|---|---|---|---|
| A | Text (haiku) | VED feature validity | Statistical stability, correlation structure, and seasonal consistency of BERT-based VED features, and whether they form an interpretable affective space for haiku. |
| B | Text (haiku) | Model comparison | TF-IDF, Word2Vec, and BERT as alternative text encoders; differences in cluster structure and seasonal trends; justification for adopting BERT as the baseline. |
| C | Music | Musical affective space (VET) | Construction and verification of Valence–Energy–Tempo features for Japanese-style instrumental tracks, and whether they yield coherent and discriminative emotional groups. |
| D | Cross-modal (text–music) | Affective alignment | Axis-wise correlations and cluster-level consistency between haiku VED and music VET, and whether a shared low-dimensional affective space supports haiku–music matching. |
| E | Multimodal + users | User perception | Participants' ratings of haiku–image–music presentations, perceived coherence, unity, and usefulness of the multimodal appreciation interface. |

### *4.1. Experiment A: Validity and Usefulness of Haiku VED Features*

This experiment verifies the validity and usefulness of the haiku emotional feature representation (VED) derived from the BERT-based intelligent data analysis.

#### 4.1.1. Experiment A-1: Validity Analysis of Extracted Features

To examine the reliability and linguistic validity of the emotional features extracted from haiku text embeddings, we conducted statistical and semantic analyses on three normalized indices: *Valence* ($V_t$), *Energy* ($E_t$), and *Dynamism* ($D_t$). The dataset consisted of 1067 haiku poems from bashoDB (Bashō Haiku Database), and all features were derived from sentence-level BERT embeddings and normalized into the $[0, 1]$ range.

(1) Descriptive Statistics

Table 2 summarizes the mean and standard deviation of each feature. All three indices were distributed within the $[0, 1]$ range, with $V_t = 0.49$, $E_t = 0.58$, and $D_t = 0.50$ on average. The balanced means and moderate standard deviations ($\sigma \approx 0.15$ to $0.19$) indicate that the extraction process yields stable and well-scaled affective representations. The histogram in Figure 3 further confirms that each dimension follows an approximately bell-shaped distribution without strong skewness or outliers. These tendencies suggest that the extracted affective space is statistically stable and unbiased.

**Table 2.** Basic statistics of normalized *Valence*, *Energy*, and *Dynamism* features.

| Feature | Mean | Standard Deviation |
|---|---|---|
| ***Valence*** **($V_t$)** | 0.49 | 0.18 |
| ***Energy*** **($E_t$)** | 0.58 | 0.15 |
| ***Dynamism*** **($D_t$)** | 0.50 | 0.19 |

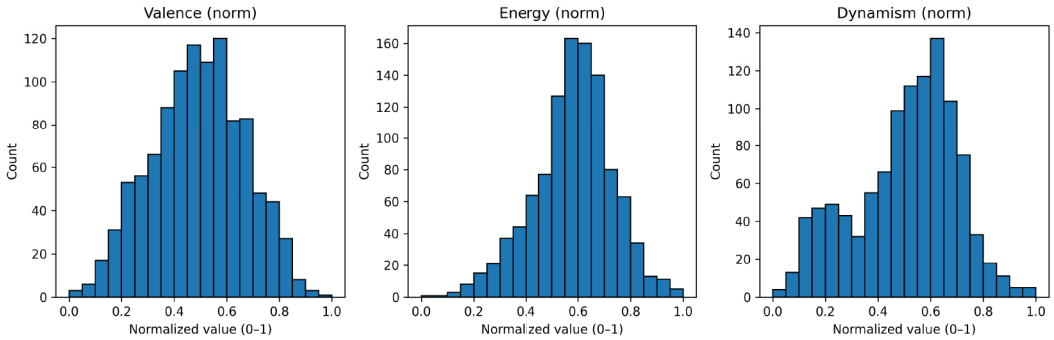



**Figure 3.** Distribution of normalized *Valence*, *Energy*, and *Dynamism* features. The balanced bell-shaped distributions indicate numerical stability and unbiased extraction.

From a linguistic perspective, the slightly lower mean of *Valence* suggests the calm and introspective tone typical of Bashō's haiku. Meanwhile, the moderately higher *Energy* and *Dynamism* values imply subtle rhythmic liveliness and emotional fluctuation within concise expressions. This balance between stillness and movement aligns with the aesthetic principles of *wabi-sabi*, supporting the contextual validity of the extracted affective dimensions.

(2) Inter-Feature Correlation

To investigate how the three affective dimensions are related to each other, we computed pairwise Spearman correlation coefficients for all 1067 haiku. The results, summarized in Table 3 and visualized as a heatmap in Figure 4, reveal a characteristic structure.

**Table 3.** Spearman correlation coefficients among normalized *Valence*, *Energy*, and *Dynamism* features.

| Feature | ***Valence*** $(V_t)$ | ***Energy*** $(E_t)$ | ***Dynamism*** $(D_t)$ |
|---|---|---|---|
| ***Valence*** $(V_t)$ | 1.00 | −0.86 | −0.71 |
| ***Energy*** $(E_t)$ | −0.86 | 1.00 | 0.51 |
| ***Dynamism*** $(D_t)$ | −0.71 | 0.51 | 1.00 |

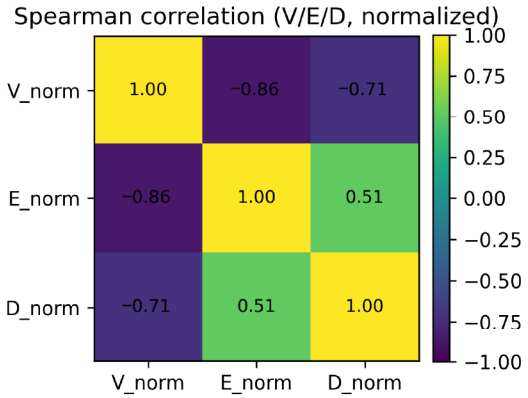



**Figure 4.** Spearman correlation heatmap among normalized *Valence*, *Energy*, and *Dynamism* features. *Valence* is strongly negatively correlated with *Energy* and *Dynamism*, while *Energy* and *Dynamism* are moderately positively correlated, indicating a structured and interpretable affective space rather than independent dimensions.

*Valence* shows strong negative correlations with both *Energy* ($\rho =$ 0.86) and *Dynamism* ($\rho = -0.71$), whereas *Energy* and *Dynamism* are moderately positively correlated ($\rho = 0.51$). This pattern suggests that haiku with calm and bright emotional tones (i.e., higher *Valence*) tend to exhibit lower intensity and less overt movement, while poems expressing tension or dramatic change often have higher *Energy* and *Dynamism* but lower *Valence*. In contrast, the positive correlation between *Energy* and *Dynamism* indicates that highly energetic haiku frequently contain explicit motion or sound imagery.

These relationships imply that the three indices are not redundant, but rather form a structured affective space: *Valence* primarily captures emotional polarity, whereas *Energy* and *Dynamism* jointly describe affective activation and expressive motion. The observed dependency is consistent with the known interplay between polarity and arousal in affective models, supporting the interpretability of the proposed VED representation for haiku.

(3) Relationship with BERT Principal Components

To determine if the extracted affective dimensions align with the intrinsic semantic structure of the original BERT embeddings, we calculated Spearman correlations between the normalized VED features and the first five principal components (PC1 to PC5) derived from the 768-dimensional embedding space. The results are shown in Table 4.

**Table 4.** Spearman correlations between normalized VED features and BERT principal components.

| Feature | PC1 | PC2 | PC3 | PC4 | PC5 |
|---|---|---|---|---|---|
| ***Valence*** $(V_t)$ | −0.77 | 0.06 | 0.29 | −0.17 | 0.17 |
| ***Energy*** $(E_t)$ | 0.53 | 0.03 | −0.26 | 0.11 | −0.16 |
| ***Dynamism*** $(D_t)$ | 0.82 | −0.11 | −0.04 | 0.11 | −0.04 |

*Valence* $(V_t)$ exhibits a strong negative correlation with the first principal component (PC1, $\rho = -0.77$), whereas *Energy* $(E_t)$ and *Dynamism* $(D_t)$ are positively correlated with PC1 ($\rho = 0.53$ and $\rho = 0.82$, respectively). The remaining components (PC2 to PC5) show only weak associations ($|\rho| < 0.3$) with all three features. This pattern suggests that PC1

primarily captures an emotional polarity axis opposing calmness to intensity, consistent with the semantic interpretation of *Valence* versus *Energy*/*Dynamism*.

These results suggest that the proposed VED features are not arbitrary projections, but rather meaningful reductions of the embedding manifold that reflect the affective organization already encoded within the pretrained BERT space. In other words, the VED representation inherits the global semantic topology of BERT while emphasizing emotion-related variance crucial to haiku expression.

(4) Seasonal Consistency Analysis

To evaluate the external validity of the extracted affective features, we examined whether *Valence*, *Energy*, and *Dynamism* vary consistently across the six seasonal categories derived from kigo (seasonal words): Spring, Summer, Autumn, Winter, New Year, and All Year. A Kruskal–Wallis non-parametric test was conducted on the normalized feature values, and the results are summarized in Table 5.

**Table 5.** Kruskal–Wallis test results for seasonal differences in normalized *Valence*, *Energy*, and *Dynamism* features. $p$-values are rounded to two decimals; for *Valence* and *Dynamism*, $p < 0.01$, and for *Energy*, $p < 0.05$. Statistical significance is reported qualitatively using the descriptors "significant" ($p < 0.05$) and "highly significant" ($p < 0.01$).

| Feature | Statistic ($H$) | $p$-Value | Significance |
|---|---|---|---|
| ***Valence*** **($V_t$)** | 20.80 | 0.00 | Highly significant |
| ***Energy*** **($E_t$)** | 11.36 | 0.04 | Significant |
| ***Dynamism*** **($D_t$)** | 18.62 | 0.00 | Highly significant |

All three indices showed statistically significant seasonal differences. *Valence* displayed the strongest effect ($H = 20.80$, $p < 0.01$), indicating that emotional polarity is highly sensitive to seasonal context. *Energy* also varied significantly across seasons ($H = 11.36$, $p < 0.05$), and *Dynamism* exhibited a clear seasonal dependency as well ($H = 18.62$, $p < 0.01$).

As illustrated in Figure 5, haiku associated with Spring and Summer show higher *Valence* and *Energy*, reflecting brightness, liveliness, and open natural imagery, whereas Autumn and Winter haiku display lower values and calmer, introspective moods. New Year haiku, treated as a distinct category in Japanese poetics, exhibit intermediate affective levels between Winter and Spring, consistent with their symbolic role as a transitional and hopeful season. These seasonal trends confirm that the extracted affective indices capture meaningful semantic differences encoded in kigo, thereby supporting the linguistic and contextual validity of the proposed VED representation.

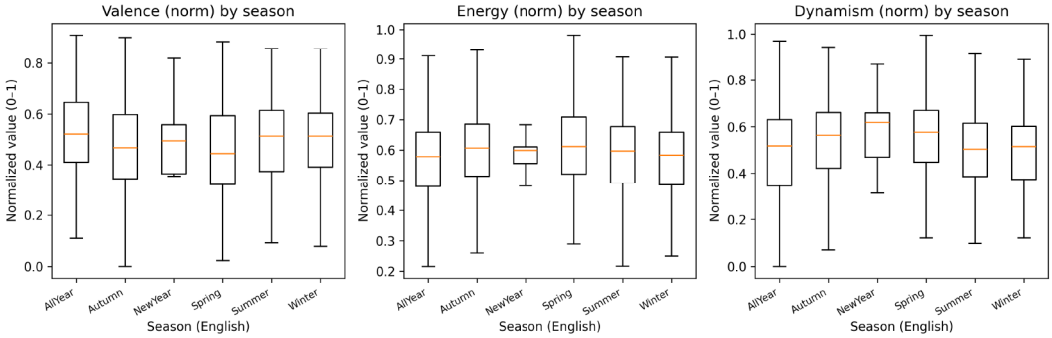



**Figure 5.** Seasonal variation of normalized *Valence*, *Energy*, and *Dynamism* features. Each boxplot represents one of six kigo categories: Spring, Summer, Autumn, Winter, New Year, and All Year. Higher *Valence* and *Energy* appear in Spring and Summer, while New Year haiku show intermediate levels between Winter and Spring.

(5) Summary

Across statistical, semantic, and linguistic perspectives, the $V_t / E_t / D_t$ representation demonstrated consistent validity: the distributions are well-scaled, inter-feature correlations confirm a structured space, and seasonal variation aligns with haiku's poetic conventions. Therefore, the proposed features are not only numerically reliable but also linguistically interpretable, providing a robust basis for subsequent clustering and cross-modal experiments.

#### 4.1.2. Experiment A-2: Usefulness Analysis of VED Features

Building upon the validity established in Experiment A-1, this experiment investigates whether the VED affective dimensions can functionally distinguish haiku with different emotional tendencies. We evaluated the functional utility of the features through unsupervised clustering and visual analysis.

(1) Clustering Procedure

The normalized *Valence*, *Energy*, and *Dynamism* values of 1067 haiku were standardized using $z$-score normalization and subjected to $k$-means clustering. To determine the optimal number of clusters, silhouette coefficients were computed for $k = 2$ to 8, as shown in Figure 6. The highest silhouette score (0.45) was obtained when $k = 2$, indicating that a two-cluster configuration provides the best balance between compactness and separation.

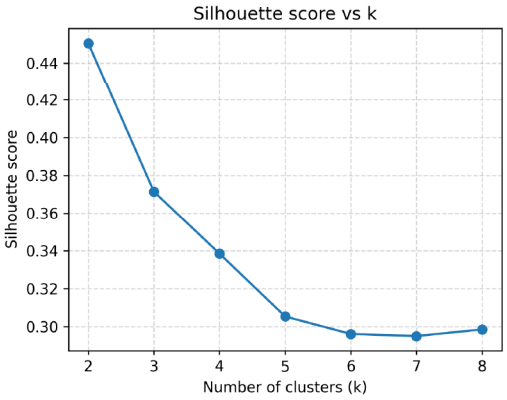



**Figure 6.** Silhouette scores for different cluster numbers ($k = 2$ to 8) using $k$-means clustering.

To strengthen methodological validation, we additionally applied Ward hierarchical clustering to the same VED feature set. As shown in Figure 7, the silhouette analysis similarly identified $k = 2$ as the optimal configuration (silhouette = 0.26). To further examine cross-method consistency, the Adjusted Rand Index (ARI) between the $k$-means and Ward clustering solutions (both at $k = 2$) was computed, yielding ARI = 0.56, which indicates moderate-to-strong agreement between the two approaches.

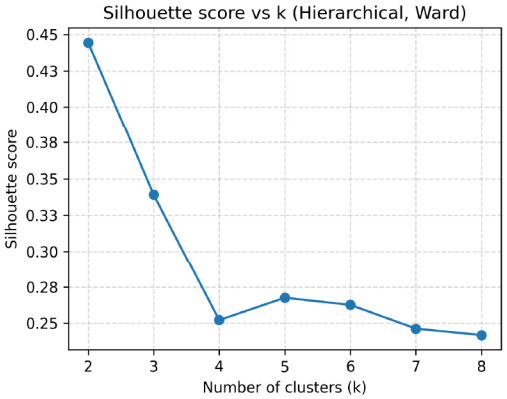



**Figure 7.** Silhouette scores for different cluster numbers ($k = 2$ to 8) using Ward hierarchical clustering.

The standardized cluster centroids are listed in Table 6. Cluster 0 shows high *Valence* but low *Energy* and *Dynamism*, representing calm, positive, and introspective haiku. Cluster 1 exhibits the opposite tendency (low *Valence* and high *Energy*/*Dynamism*), corresponding to dramatic or high-tension haiku often depicting movement, sound, or contrast.

**Table 6.** Cluster centers for *Valence*, *Energy*, and *Dynamism* features (standardized $z$-scores and normalized $[0, 1]$ values).

| Cluster | $z$-Score | | | Normalized [0,1] | | |
|---|---|---|---|---|---|---|
| | $V_z$ | $E_z$ | $D_z$ | $V_{01}$ | $E_{01}$ | $D_{01}$ |
| 0 (Calm/Positive) | 1.01 | −0.94 | −0.96 | 0.67 | 0.44 | 0.32 |
| 1 (Dramatic/Active) | −0.58 | 0.53 | 0.55 | 0.39 | 0.66 | 0.61 |

Based on these centroids, the two clusters can be characterized as follows:

- **Cluster 0 (Calm/Positive):** High *Valence* and low *Energy*/*Dynamism*. This group represents calm, gentle, and observational haiku that convey a predominantly positive emotional tone.
- **Cluster 1 (Dramatic/Active):** Low *Valence* and high *Energy*/*Dynamism*. This group corresponds to dramatic or high-intensity haiku that evoke movement, tension, sound imagery, or vivid sensory contrast.

(2) Visualization and Interpretation

Figure 8 visualizes the haiku distribution in the $V_t/E_t/D_t$ feature space. The two clusters are clearly separated: Cluster 0 (blue) occupies the region of high *Valence* and low *Energy*/*Dynamism*, while Cluster 1 (orange) lies in the opposite region. This separation suggests that emotional polarity and expressive intensity form two complementary axes in haiku affective representation.

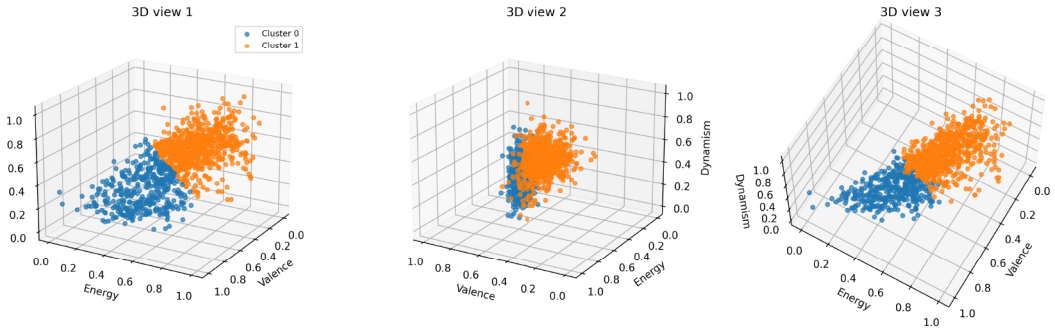



**Figure 8.** 3-D visualization of haiku distribution in the VED feature space. Colors denote $k$-means cluster assignments ($k = 2$). Cluster 0 corresponds to calm/positive haiku, and Cluster 1 to dramatic/active haiku with high expressive intensity.

The 2-D projections in Figure 9 show that cluster boundaries align naturally with emotional polarity and arousal: haiku with high *Valence* and low *Energy* correspond to serenity and observation, whereas those with low *Valence* and high *Energy* capture movement, tension, or sound imagery.

Taken together, the 3-D and 2-D scatterplots visually confirm the numerical tendencies already indicated by the $z$-score centroids. Cluster 0 occupies the region of higher *Valence* and lower *Energy*/*Dynamism*, consistent with its centroid profile. It corresponds to calm, gentle, and observant haiku with a predominantly positive emotional tone. In contrast, Cluster 1 appears in the region of lower *Valence* and higher *Energy*/*Dynamism*, matching its

centroid pattern and representing haiku that evoke movement, tension, sound, or vivid sensory intensity.

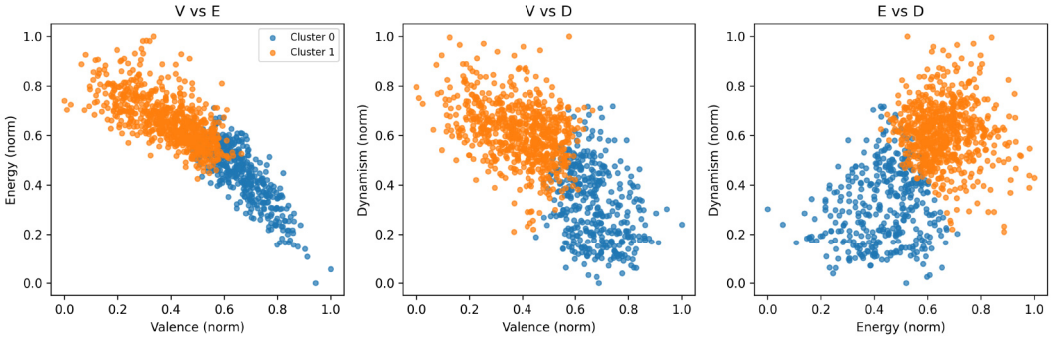



**Figure 9.** 2-D projections of the haiku VED feature space. The separation between clusters reflects affective differences in polarity (*Valence*) and intensity (*Energy*/*Dynamism*).

These distributions indicate that the VED feature space captures a meaningful axis of emotional polarity and expressive intensity rather than noise. Although the cluster boundaries exhibit smooth transitions rather than sharp separations, the overall affective tendencies remain coherent and semantically interpretable.

Furthermore, although not visualized here, the alternative Ward hierarchical clustering produced the same optimal cluster number ($k = 2$) and showed moderate agreement with the $k$-means solution (ARI = 0.56), suggesting that both algorithms uncover a comparable underlying contrast between calm/positive haiku and dramatic, high-intensity haiku.

(3) Summary

The results of Experiment A-2 demonstrate that the proposed VED representation is not only statistically valid but also functionally useful. Through unsupervised clustering, the system automatically divided haiku into two semantically interpretable groups (calm/positive and dramatic/active). The centroid differences exceed one standard deviation across all dimensions, and the clustering achieved a mean silhouette coefficient of 0.45, confirming that the observed separation is both statistically stable and semantically interpretable.

These findings indicate that the VED space captures meaningful emotional structure rather than arbitrary numeric variation, providing a reliable foundation for subsequent cross-modal alignment and evaluation in Experiment D.

### *4.2. Experiment B: Comparative Analysis of Alternative Text Representations*

Following the verification of the VED feature validity in Experiment A, Experiment B investigates the suitability of alternative text representation models for extracting affective features from haiku. Specifically, three text representation approaches (TF–IDF, Word2Vec, and BERT) were examined. This experiment aims to justify our choice of BERT by examining how different text representations influence the construction of the affective feature space and evaluating the relative stability and interpretability of each approach.

All three encoders were evaluated under an identical pipeline consisting of sentence embedding, computation of VED statistics (mean, standard deviation, and maximum), min–max normalization to the $[0, 1]$ range, $z$-score standardization, and $k$-means clustering with the number of clusters varied from 2 to 8. All intermediate results and visualizations were documented to ensure transparency and reproducibility.

(1) Clustering Tendency and Silhouette Analysis

To compare the separability of affective feature spaces obtained from different text representations, the silhouette coefficients of TF–IDF, Word2Vec, and BERT embeddings were computed for cluster numbers $k = 2$ to 8 (see Figure 10). As summarized in Table 7, TF–IDF achieved the highest silhouette coefficient (0.60 at $k = 2$), followed by BERT (0.45) and Word2Vec (0.30), indicating that TF-IDF forms the most geometrically compact clusters in the VED space.

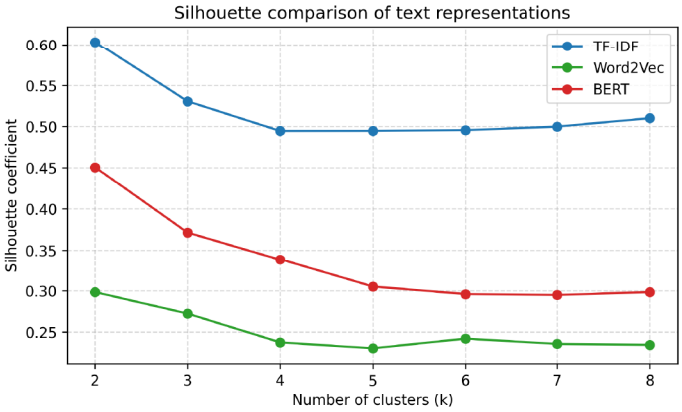



**Figure 10.** Comparison of silhouette coefficients across TF-IDF, Word2Vec, and BERT embeddings ($k = 2$ to 8).

**Table 7.** Average silhouette coefficients for each encoder ($k = 2$ to 8).

| Model | $k = 2$ | $k = 3$ | $k = 4$ | $k = 5$ | $k = 6$ | $k = 7$ | $k = 8$ |
|---|---|---|---|---|---|---|---|
| **TF-IDF** | 0.60 | 0.53 | 0.49 | 0.49 | 0.50 | 0.50 | 0.51 |
| **Word2Vec** | 0.30 | 0.27 | 0.24 | 0.23 | 0.24 | 0.24 | 0.23 |
| **BERT** | 0.45 | 0.37 | 0.34 | 0.31 | 0.30 | 0.30 | 0.30 |

At first glance, these results suggest that TF-IDF is the most effective representation for clustering haiku embeddings. However, the silhouette coefficient primarily reflects geometric separability rather than semantic validity. TF-IDF's high score indicates clear boundaries between clusters, but such separation likely arises from lexical frequency patterns rather than underlying affective semantics. In contrast, BERT produced moderate silhouette values that decrease gradually with larger $k$, indicating a structured yet non-fragmented feature space where emotional expressions are distributed more continuously. Word2Vec, on the other hand, yielded low and unstable scores, suggesting that its embeddings lack meaningful differentiation for short poetic texts. Therefore, although TF-IDF numerically outperforms the other models, its advantage may not reflect accurate emotional organization.

(2) Feature-Space Visualization by PCA

To visualize and compare the internal distribution of haiku embeddings in the affective feature space, we applied principal component analysis (PCA) to the normalized VED features derived from each model. As shown in Figure 11, the TF-IDF representation produced a highly compressed structure along the first principal component (PC1 = 84.5%), indicating that most variance originates from a single lexical dimension. Word2Vec formed a roughly isotropic cluster (PC1 = 49.7%, PC2 = 37.4%), suggesting a lack of dominant affective axes. By contrast, BERT generated an elongated yet structured distribution (PC1 = 79.7%, PC2 = 16.8%), suggesting the presence of multiple, interpretable emotional gradients.

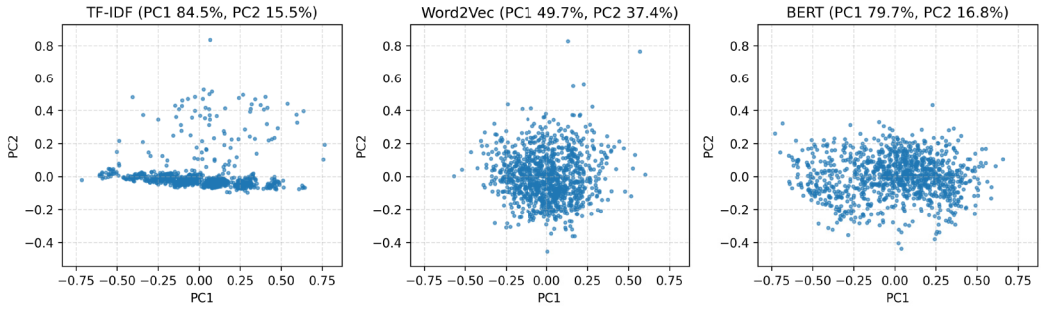



**Figure 11.** Two-dimensional PCA visualization of haiku embeddings based on TF-IDF (**left**), Word2Vec (**middle**), and BERT (**right**). Each point represents one haiku.

(3) Cluster-Level Comparison of VED Distributions

To further investigate the affective interpretability of the clusters obtained in Section 4.2, the normalized values were compared across the two primary clusters ($k = 2$), as shown in Figure 12. For the BERT encoder, clear contrasts emerged between clusters: Cluster 0 exhibited high *Valence* and low *Energy*/*Dynamism* (calm expressions), whereas Cluster 1 displayed the opposite (dynamic imagery). TF-IDF and Word2Vec, in contrast, showed weaker and less meaningful separations, often dominated by a single dimension (typically *Energy*). These results suggest that the BERT-based affective space provides more semantically coherent and emotionally interpretable clusters.

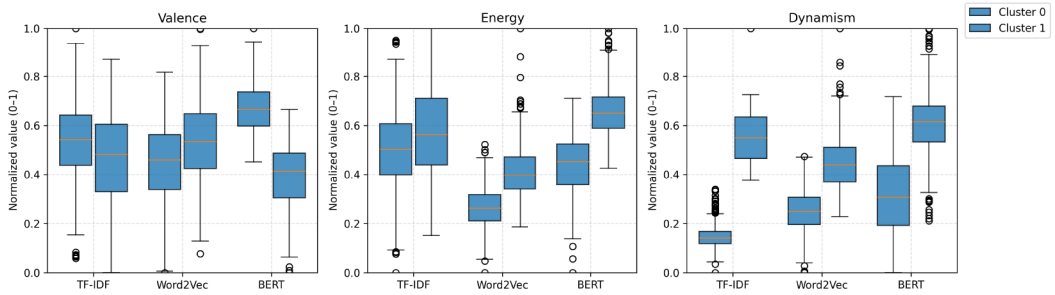



**Figure 12.** Distribution of normalized *Valence*, *Energy*, and *Dynamism* across two clusters ($k = 2$) for each model.

(4) Seasonal Trend Analysis of Affective Indices

To explore whether the extracted VED features capture the intrinsic seasonality of haiku, the mean values of $V_t$, $E_t$, and $D_t$ were calculated for six seasonal categories: Spring, Summer, Autumn, Winter, New Year, and All Year. Figure 13 compares the seasonal averages of TF-IDF and BERT. BERT exhibited gentle and consistent seasonal variation (higher *Valence* and *Dynamism* in Summer, and lower in Winter), which aligns with conventional affective impressions in Japanese seasonal poetry. TF-IDF, by contrast, showed irregular or flattened patterns, implying a weaker sensitivity to the latent emotional tone of seasonal expressions.

(5) Summary

Overall, although TF-IDF achieved the highest numerical silhouette coefficient, its PCA and seasonal analyses revealed structural bias and limited interpretability of affect. Word2Vec captured general lexical relations but failed to differentiate emotional subtleties. BERT provided the most balanced representation: its clusters corresponded to meaningful emotional contrasts, and its seasonal variations were consistent with cultural expectations in haiku. Therefore, BERT was confirmed as the most suitable encoder for our intelligent data analysis pipeline and was adopted for all subsequent cross-modal experiments (Experiments C and D).

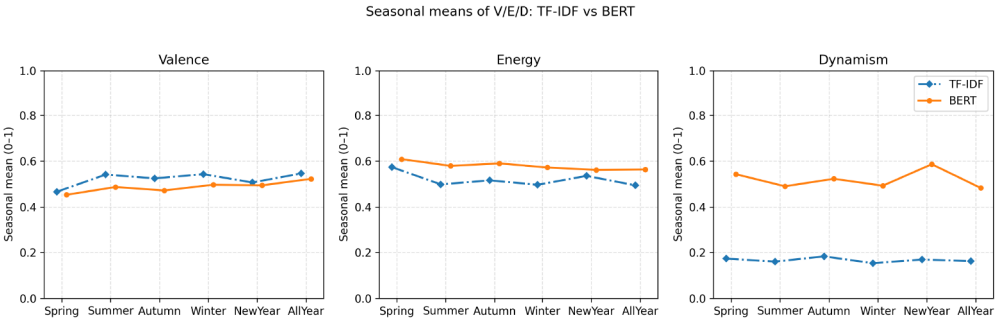



**Figure 13.** Seasonal means of *Valence*, *Energy*, and *Dynamism* for TF-IDF and BERT encoders.

### *4.3. Experiment C: Verification of the Musical Affective Feature Space*

Following the validation of text-based affective features in Experiments A and B, this experiment shifts to the auditory domain. Its purpose is to examine whether the musical indices (*Valence* ($V_m$), *Energy* ($E_m$), and *Tempo* ($T_m$)) can also represent a coherent and interpretable affective space. We implemented the analysis using a unified data analysis pipeline that combines acoustic feature extraction, robust normalization, statistical verification, and unsupervised clustering. This pipeline explores the underlying emotional structure of instrumental music.

(1) Feature Extraction

Each instrumental piece from the Free BGM DOVA-SYNDROME dataset was resampled to 22,050 Hz, trimmed for silence, and analyzed using `librosa` feature extraction functions, many of which are built upon the short-time Fourier transform (STFT). A set of acoustic descriptors was extracted, including tempo (BPM), RMS Energy, zero-crossing rate, spectral centroid, bandwidth, roll-off, spectral contrast, chroma features, 13-dimensional MFCCs, and tonnetz. The mean and standard deviation of each descriptor were calculated and stored, resulting in a unified feature table for all tracks, which served as the foundation for subsequent normalization and clustering analyses.

In contrast to previous chroma-based implementations [29], the present study redefines *Valence* as a weighted measure of spectral brightness, as defined in Equation (5) in Section 3.3. This formulation approximates the perceived brightness of an audio signal. Perceived brightness is commonly associated with positive emotions when listening to music. This formulation also provides greater numerical stability. *Energy* was defined as the mean RMS amplitude, and *Tempo* as the estimated beats per minute (BPM).

(2) Normalization and Statistical Validation

To ensure scale independence and suppress outliers, the range of each raw feature was clipped to the 1st–99th percentiles and normalized to the range of $[0, 1]$ using min–max scaling. The resulting indices ($V_{\text{norm}}$, $E_{\text{norm}}$, and $T_{\text{norm}}$) were then standardized by $z$-score before multivariate analysis, as illustrated in Figure 14.

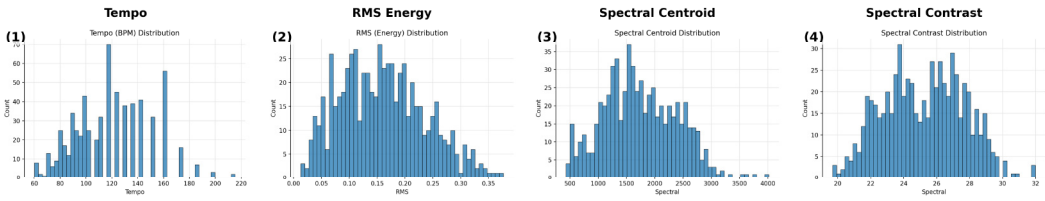



**Figure 14.** Overall distributions of four representative musical descriptors after robust normalization: (**1**) Tempo (BPM), (**2**) RMS Energy, (**3**) Spectral Centroid (Hz), and (**4**) Spectral Contrast (dB). Each feature was clipped to the 1st–99th percentile and rescaled to the $[0, 1]$ range. The distributions are smooth and balanced, indicating numerically stable feature extraction.

A Spearman correlation heatmap (Figure 15, Left) revealed strong positive correlations among tempo, RMS energy, and spectral brightness ($r \approx 0.8$). This indicates that these descriptors possess high internal consistency in describing the intensity and activity level of the music.

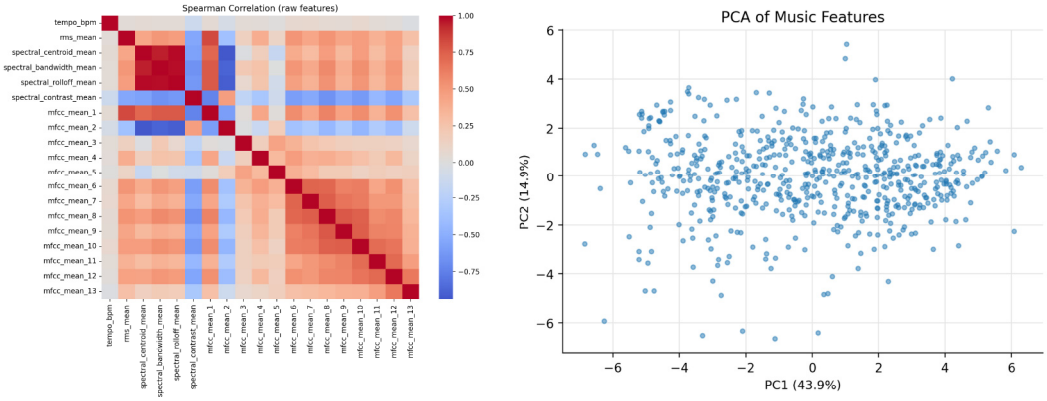



**Figure 15.** (**Left**) Spearman correlation heatmap among acoustic features, showing high internal consistency. (**Right**) PCA scatter plot, where PC1 corresponds to the Energy–Brightness axis and PC2 to the Tempo–Contrast axis, explaining about 67% of total variance.

Furthermore, the PCA scatter plot (Figure 15, Right) demonstrates that the feature space is well-structured. Two principal components (PC1, corresponding to the Energy–Brightness axis, and PC2, corresponding to the Tempo–Contrast axis) explain approximately 67% of the total variance.

These findings together validate that the extracted features capture perceptually meaningful dimensions of musical emotion, forming a robust basis for the subsequent clustering analysis.

(3) Clustering Analysis

To examine the discriminative power of the VET features, the standardized $[V, E, T]$ vectors of 643 tracks were grouped using $k$-means clustering ($k$ varied from 2 to 8). As shown in Figure 16, the silhouette coefficient was computed for each $k$ to evaluate cluster compactness and separation.

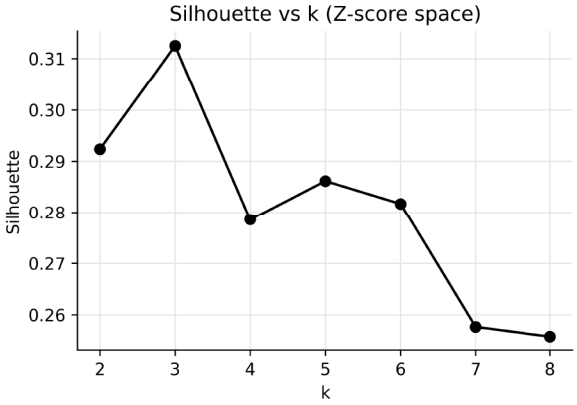



**Figure 16.** Silhouette scores for different cluster numbers ($k = 2$ to 8) using $k$-means clustering. The highest silhouette value (0.31) is obtained at $k = 3$, indicating that a three-cluster configuration best represents the underlying emotional structure.

To strengthen methodological robustness, we additionally applied Ward hierarchical clustering to the same VET feature space. As shown in Figure 17, the silhouette analysis again selected $k = 3$ as the optimal cluster number. To examine the consistency between

the two clustering approaches, the Adjusted Rand Index (ARI) comparing the $k$-means and Ward clustering solutions (both at $k = 3$) was computed, resulting in ARI = 0.50. This indicates moderate agreement and suggests that both methods capture a comparable underlying emotional structure in the musical feature space.

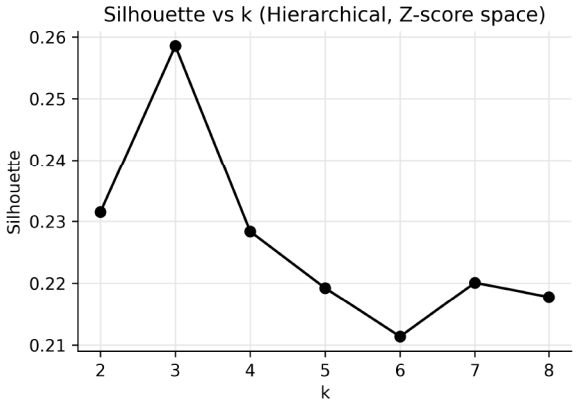



**Figure 17.** Silhouette scores for different cluster numbers ($k = 2$ to 8) using Ward hierarchical clustering. The highest silhouette value is obtained at $k = 3$, indicating that a three-cluster configuration provides the best structure under this method.

As shown in Figure 18, three characteristic clusters were observed:

- **Cluster 0:** High *Valence*, high *Energy*, moderate *Tempo* — bright, lively, and uplifting pieces.
- **Cluster 1:** Low *Valence*, low *Energy*, high *Tempo* — dark but dynamic pieces with strong rhythm.
- **Cluster 2:** Low *Valence*, low *Energy*, slow *Tempo* — calm, soft, and introspective pieces.

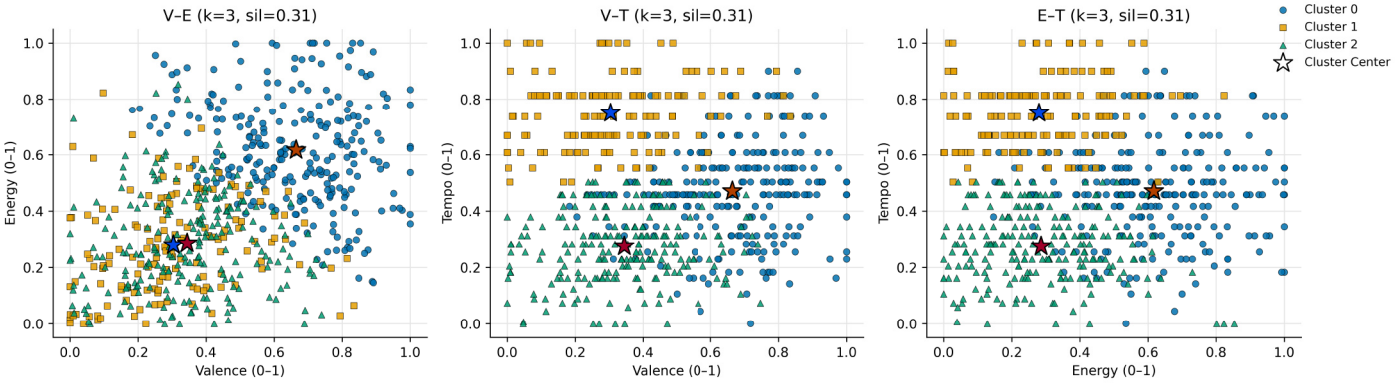



**Figure 18.** Clustering results projected onto *Valence–Energy*, *Valence–Tempo*, and *Energy–Tempo* planes. Each point represents one music piece; color indicates cluster membership ($k = 3$, silhouette = 0.31).

The normalized centroids were approximately (0.80, 0.82, 0.35) for Cluster 0, (0.30, 0.28, 0.75) for Cluster 1, and (0.40, 0.35, 0.15) for Cluster 2, respectively.

Although the cluster boundaries are continuous rather than sharply separated, they correspond well to intuitive affective categories and reflect smooth transitions among musical styles.

Taken together, the $k$-means and Ward clustering results demonstrate that the VET features support a stable three-cluster affective structure. The three groups are bright/lively (Cluster 0), dark/dynamic (Cluster 1), and calm/soft (Cluster 2). These groups emerge consistently across algorithms, confirming that the musical affective space reflects coherent and semantically interpretable emotional distinctions rather than method-specific artifacts.

(4) Summary

Experiment C demonstrated that the proposed musical affective indices (VET) are both statistically valid and perceptually interpretable. Robust normalization produced stable, evenly distributed features, while correlation and PCA analyses confirmed the internal coherence of the *Valence*, *Energy*, and *Tempo* dimensions. Unsupervised clustering revealed three emotionally consistent groups that naturally correspond to variations in brightness, intensity, and activity. These results validate our intelligent data analysis pipeline for music and establish the VET representation as a reliable low-dimensional affective space for modeling musical emotion. It forms the auditory basis for the cross-modal haiku–music alignment analysis presented in Experiment D.

### *4.4. Experiment D: Cross-Modal Alignment Analysis Between Haiku and Music*

Building on the established affective representations of haiku (Experiments A and B) and instrumental music (Experiment C), Experiment D evaluates whether the two modalities share a consistent affective structure. This experiment examines cross-modal alignment in two complementary analyses:

- **Axis-wise alignment:** Linear correspondence between the affective dimensions of haiku (VED) and music (VET).
- **Cluster-level alignment:** Global geometric similarity between haiku clusters and music clusters, and the consistency of actual haiku–music matches.

Experiment D-1 produced matching pairs that served as the basis for Experiment D-2, which used the clustering results from Experiments A and C.

#### 4.4.1. Experiment D-1: Axis-Wise Alignment

For each haiku feature vector,

$$h_{\mathrm{VED}} = (V_t, E_t, D_t)$$

the Euclidean distance to all 643 music vectors

$$m_{\mathrm{VET}} = (V_m, E_m, T_m)$$

was computed within the shared three-dimensional affective space $(V, E, D/T)$. The nearest music track was then selected as the system's predicted background music, resulting in 1067 haiku–music pairs with affective values $(V_t, E_t, D_t; V_m, E_m, T_m)$.

To evaluate axis-wise consistency, Pearson correlation coefficients were computed for

$$(V_t, V_m), \quad (E_t, E_m), \quad (D_t, T_m)$$

For each axis, 1000 bootstrap resamples were used to estimate 95% confidence intervals, and a random baseline was obtained by repeatedly shuffling the music-side values while keeping the haiku features fixed.

As shown in Figure 19, the scatter plots exhibit tightly concentrated diagonal patterns, indicating strong axis-wise alignment between haiku features and their matched music counterparts.

Table 8 reports the corresponding correlation coefficients, confidence intervals, and the random shuffled baselines. The empirical correlations for all three axes are substantially and significantly higher than the baseline values, indicating that the matches preserve a strong axis-wise correspondence in the shared affective space, well above what would be expected by random pairing.

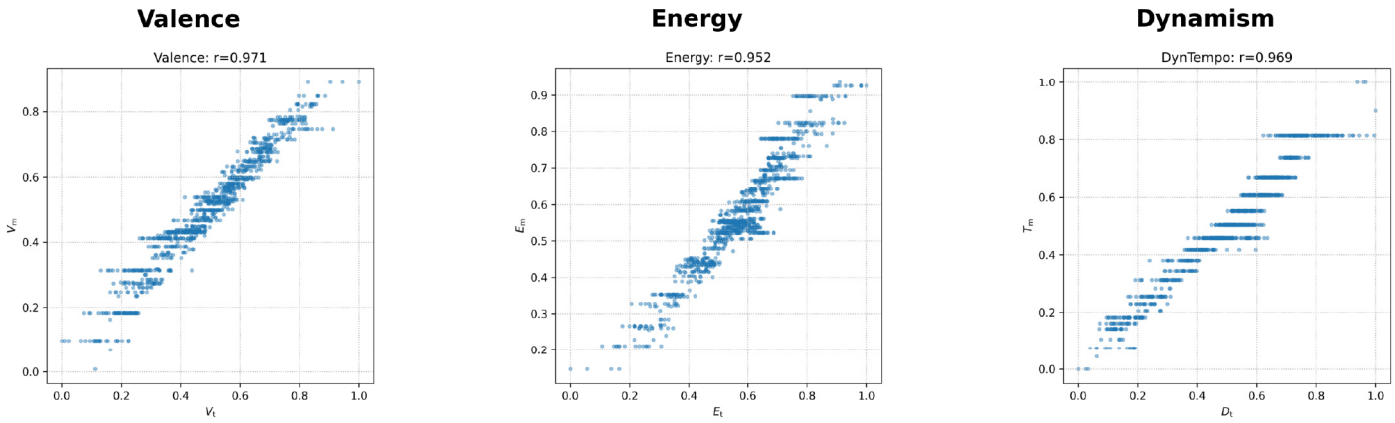



**Figure 19.** Axis-wise haiku–music alignment (Experiment D-1). Scatter plots for *Valence*, *Energy*, and *Dynamism/Tempo* show diagonal tendencies, indicating high cross-modal consistency. The displayed $r$ values correspond to the empirical correlations of the matched pairs.

**Table 8.** Axis-wise correlation between haiku (VED) and music (VET) features in Experiment D-1. Empirical correlations are contrasted with the randomly shuffled baseline.

| Axis | Correlation $r$ | 95% CI | Baseline Mean | Baseline SD |
|---|---|---|---|---|
| ***Valence*** | 0.971 | [0.968, 0.974] | 0.000 | 0.035 |
| ***Energy*** | 0.952 | [0.946, 0.957] | 0.003 | 0.030 |
| ***Dynamism/Tempo*** | 0.969 | [0.965, 0.973] | 0.001 | 0.034 |

#### 4.4.2. Experiment D-2: Cluster-Level Alignment

While Experiment D-1 assessed cross-modal alignment at the level of individual affective axes, Experiment D-2 investigates whether haiku and music share a consistent structure at the *cluster* level. This analysis draws on three components: (i) the two haiku clusters from Experiment A-2, (ii) the three music clusters from the analysis in Experiment C-2, and (iii) the 1067 haiku–music pairs obtained in Experiment D-1. This analysis involved two complementary steps: (1) geometric similarity between haiku and music cluster centers, and (2) cluster consistency of the actual haiku-music matches.

(1) Geometric Alignment of Cluster Centers

For each haiku cluster $H_k$ and music cluster $M_l$, we computed the Euclidean distance between their centroids in the normalized affective space $(V, E, D/T)$. Let $\mu_{H_k}$ and $\mu_{M_l}$ denote the corresponding cluster centers; the distance matrix is given by

$$d(H_k, M_l) = \|\mu_{H_k} - \mu_{M_l}\|_2$$

resulting in the following $2 \times 3$ configuration:

| | $M0$ | $M1$ | $M2$ |
|---|---|---|---|
| $H0$ | 0.235 | 0.588 | 0.363 |
| $H1$ | 0.310 | 0.415 | 0.505 |

which is visualized as a heatmap in Figure 20.

For each haiku cluster, we then took the minimum distance to the three music clusters and averaged these minimum distances across the two haiku clusters. The resulting average minimum distance was

$$\text{mean}(d_{\min}) = 0.273$$

which coincided with the random-shuffle baseline $(0.273 \pm 1.1 \times 10^{-16})$. This suggests that, although haiku and music clusters occupy comparable regions in the affective space, their global geometries are not strongly aligned by themselves. Therefore, we next examine whether the actual haiku–music matches exhibit stronger cross-modal consistency at the cluster level.

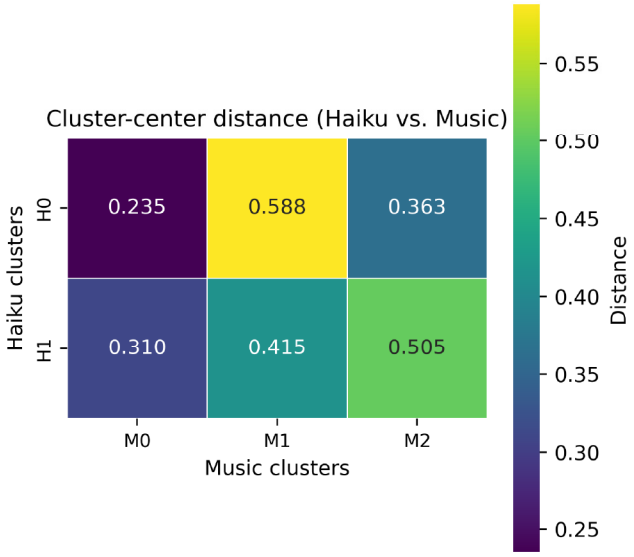



**Figure 20.** Cluster-center distances between haiku and music clusters (Experiment D-2). Each cell shows the Euclidean distance between a haiku cluster centroid $H_k$ and a music cluster centroid $M_l$ in the normalized affective space $(V, E, D/T)$. Lower values indicate closer affective positions.

(2) Cluster Consistency of Haiku–Music Matches

To evaluate whether haiku from a given cluster tend to be matched with music from specific music clusters, we constructed a contingency table between haiku-cluster labels (cluster_H) and music-cluster labels (cluster_M) for all 1067 haiku–music pairs obtained in Experiment D-1. The resulting frequency matrix is shown in Table 9.

**Table 9.** Contingency matrix between haiku clusters (cluster_H) and music clusters (cluster_M) for the 1067 haiku–music pairs generated in Experiment D-1.

| cluster_H | cluster_M = 0 | cluster_M = 1 | cluster_M = 2 |
|---|---|---|---|
| **0** | 335 | 2 | 50 |
| **1** | 509 | 170 | 1 |

Normalizing each row of Table 9 yields the conditional probability $P(c_M \mid c_H)$, i.e., the probability that a haiku in cluster $c_H$ is matched with music from cluster $c_M$. The resulting $2 \times 3$ probability matrix is

| | $M0$ | $M1$ | $M2$ |
|---|---|---|---|
| $H0$ | 0.866 | 0.005 | 0.129 |
| $H1$ | 0.749 | 0.250 | 0.001 |

and is visualized in Figure 21.

Both haiku clusters exhibit a strong tendency to be matched with music cluster M0: $P(M0 \mid H0) = 0.866$ and $P(M0 \mid H1) = 0.749$. In addition, haiku cluster H1 shows a secondary alignment toward music cluster M1 ($P(M1 \mid H1) = 0.250$), whereas matches to M2 are rare for both haiku clusters.

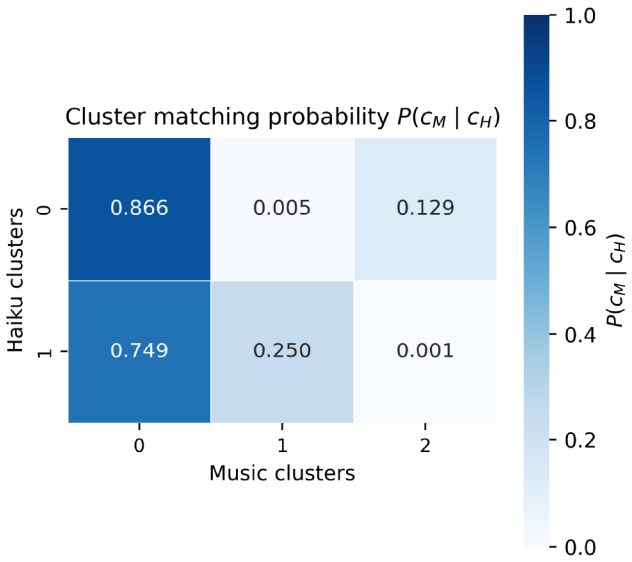



**Figure 21.** Cluster matching probabilities $P(c_M \mid c_H)$ (Experiment D-2). Each cell shows the conditional probability that a haiku from cluster $H_k$ is matched to music from cluster $M_l$, based on the 1067 haiku–music pairs. Darker cells indicate higher matching probability.

To summarize this tendency, we computed the cluster purity for each haiku cluster as

$$\text{purity}(H_k) = \max_l P(c_M = l \mid c_H = k)$$

resulting in purities of 0.866 for H0 and 0.749 for H1. The average cluster purity was therefore $\text{purity}_{\text{avg}} = 0.807$, which exceeded the randomized baseline $(0.791 \pm 0.003)$ obtained by repeatedly shuffling music-cluster labels. Figure 22 illustrates the per-cluster purity and the random baseline.

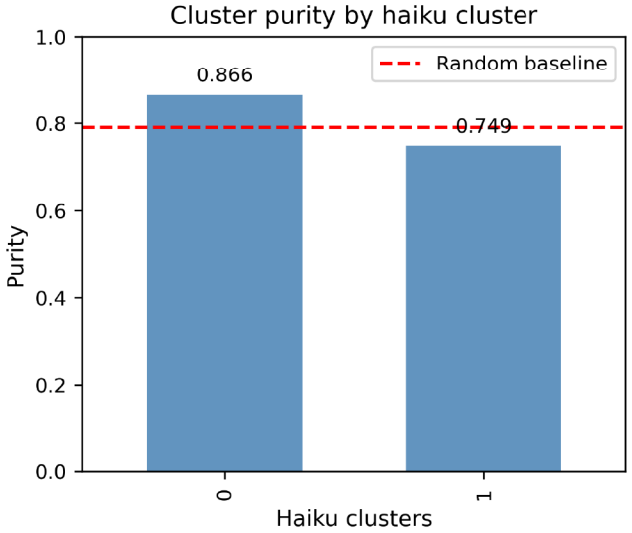



**Figure 22.** Cluster purity of haiku–music matches (Experiment D-2). Bars indicate the purity of each haiku cluster, defined as $\max_l P(c_M = l \mid c_H = k)$, and the dashed horizontal line shows the average purity under random shuffling of music-cluster labels. The observed average purity (0.807) is higher than the random baseline (0.791).

(3) Summary

Overall, Experiment D-2 shows that, although the centroids of haiku and music clusters are not strongly aligned in a purely geometric sense, the actual haiku–music matches produced by the system exhibit statistically reliable cluster-level consistency well above

random chance. Together with the axis-wise correlations observed in Experiment D-1, these results help validate the presence of a coherent cross-modal affective structure shared between haiku and instrumental music.

#### 4.4.3. Experiment D-3: Reproducibility and Stability of Image Generation

While Experiments A–D evaluate textual and musical affective structures, the multimodal system also relies on generated images. To examine whether the image-generation module produces stable and reproducible visual outputs under a fixed prompt template, we conducted a controlled reproducibility experiment.

(1) Experimental Setup

To assess the reproducibility and stability of image generation, six representative haiku—one from each of the six seasonal categories (spring, summer, autumn, winter, New Year, and all-year)—were selected for this experiment. For each haiku, the same prompt template (Section 3.4) was used across 10 independent generation trials, yielding a total of 60 images.

All haiku metadata used to populate the prompt template are provided in Appendix A. Each generation output was accepted as returned by the model without manual filtering, ensuring that the collected samples reflect the system's natural variability under fixed prompting conditions.

All images were generated using the `DALL·E (GPT-image-1)` model with a fixed resolution of $1024 \times 1024$ pixels, one image per call, default decoding settings, and no externally specified random seed. Thus, any variation across trials reflects the model's intrinsic sampling stochasticity rather than changes in user-controlled parameters.

Note that this experiment focuses solely on reproducibility; the semantic appropriateness of the generated images was evaluated separately in the user study of Experiment E.

(2) Consistency Analysis and Results

Visual reproducibility was quantified using pairwise CLIP cosine similarity scores among the 10 images generated for each haiku, computed with the OpenAI CLIP ViT-B/32 model. For each haiku, the upper-triangular values of the similarity matrix were used to compute the mean and standard deviation, representing the central tendency and stochastic variation of the generated outputs.

We report both statistics to jointly capture reproducibility and sensitivity to stochastic variation. Table 10 summarizes the results. Across all six haiku, similarity scores were consistently high (0.90–0.96), and variation across repeated generations remained small (standard deviation 0.01–0.03). These findings indicate that the fixed prompt template imposes a strong semantic constraint on the output, while residual stochastic variation primarily affects stylistic nuances rather than core visual semantics.

**Table 10.** Within-set CLIP similarity among ten generated images per haiku. Higher similarity values indicate greater reproducibility, while lower standard deviations reflect more stable generation behavior.

| Haiku ID | Mean Similarity | Std. Deviation |
|---|---|---|
| 231 | 0.93 | 0.03 |
| 234 | 0.91 | 0.03 |
| 648 | 0.94 | 0.01 |
| 78 | 0.93 | 0.02 |
| 839 | 0.94 | 0.01 |
| 932 | 0.96 | 0.01 |

(3) Summary

Across all haiku, high similarity values show that repeated generation with the same prompt produces visually consistent results. Although small stylistic fluctuations arise due to stochastic sampling, essential semantic elements remain stable. Thus, the image-generation module functions as a reliable component of the multimodal haiku appreciation system and does not introduce unwanted variability into either the cross-modal alignment analyses or the user-evaluation study in Experiment E.

### *4.5. Experiment E: User Evaluation of the Haiku Appreciation Experience*

While Experiments A to D established the validity of the affective feature space and the cross-modal matching mechanism, Experiment E investigates how users perceive the system's multimodal haiku presentations. To assess the practical effectiveness of the generated illustrations, matched background music, and their combination, a user study was conducted in which participants viewed a series of haiku accompanied by the system-produced visual and auditory outputs. After each presentation, participants rated multiple aspects of the haiku appreciation experience using a structured questionnaire.

#### 4.5.1. Method

(1) Evaluation Procedure

Each participant viewed five multimodal haiku demonstrations (D1 to D5). Each demonstration consisted of the following four components:

- the original Japanese haiku;
- a generated watercolor-style illustration;
- an affectively matched instrumental music excerpt; and
- a short TTS narration of the haiku reading.

The haiku, image, and audio sets were presented in a randomized order. Participants were instructed to read the poem, observe the illustration, listen to the narration, and listen to the background music excerpt (approximately 20 to 25 s) before providing their subjective ratings.

(2) Participants

A total of eleven students from Yamaguchi University participated in the study. The participants ranged in age from 19 to 24 years (mean age approximately 22; 8 male, 3 female) and included both Japanese ($n = 5$) and Chinese ($n = 6$) students. All participants had encountered classical Japanese haiku during their compulsory education, and none had specialized training in haiku composition, literary criticism, music theory, or image-generation technologies. Participants viewed the five demonstrations and completed a structured questionnaire after each one, providing ratings on a five-point Likert scale as well as brief free-form comments. It allowed us to collect both quantitative and qualitative feedback on the clarity, harmony, and overall usefulness of the multimodal presentations. This questionnaire did not require ethics committee approval, as it did not involve animal or human clinical trials and did not raise significant ethical concerns. This study adhered to the ethical principles outlined in the Declaration of Helsinki. All participants were informed of the questionnaire's purpose and provided their informed consent before participating. All participants were guaranteed anonymity and confidentiality, and participation was entirely voluntary.

(3) Measures

After viewing each demo, participants evaluated three aspects of the presentation using a five-point Likert scale (1 = strongly disagree, 5 = strongly agree):

- the appropriateness of the generated illustration for the haiku,
- the harmony between the haiku and the background music,
- the overall sense of unity of the harmonious multimodal presentation.

After completing all five demos, participants also provided overall impressions of the system, including perceived novelty, emotional clarity, and general usefulness for haiku appreciation, followed by optional free-form comments. This combination of repeated individual ratings and final overall feedback allowed us to assess both within-demo and across-demo user impressions.

#### 4.5.2. Results

Figure 23 illustrates the distribution of participants' ratings for the five demonstrations (D1 to D5). Across all demos, most scores range between 3 and 5, indicating generally positive impressions of the generated illustrations, the selected background music, and their combined presentation. In particular, D1, D2, D4, and D5 show a concentration of scores around 4. In contrast, D3 shows greater variability in user perception, indicating a wider range of scores.

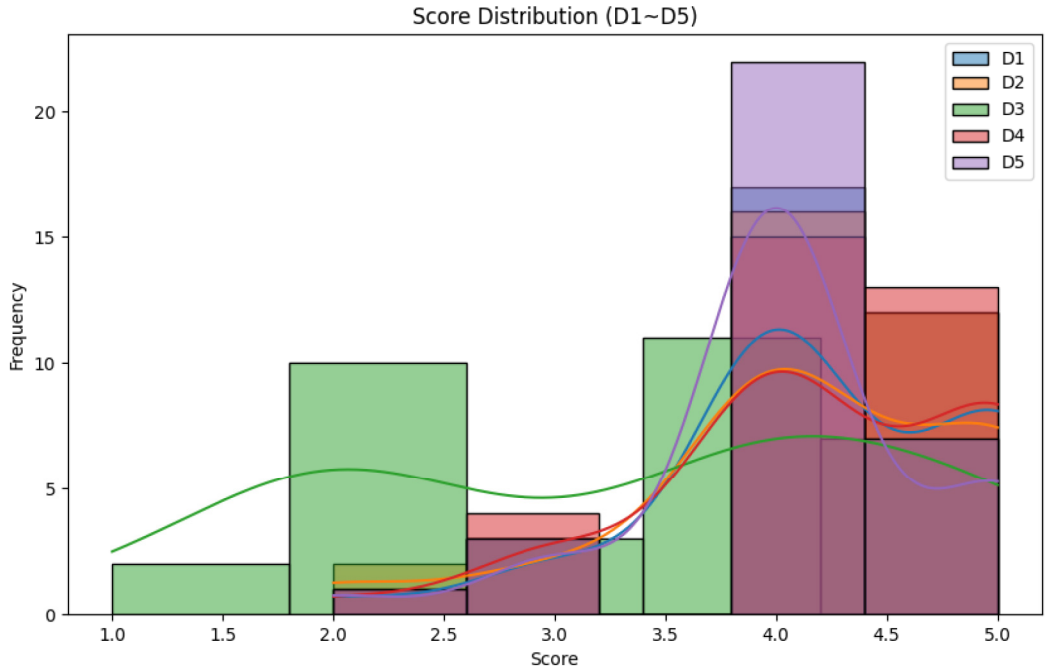



**Figure 23.** Score distribution for demonstrations (D1 to D5) is shown in the histograms and kernel density estimates. These show the distribution of participant ratings for each demonstration. Scores range from 1 (strongly disagree) to 5 (strongly agree).

Figure 24 summarizes the mean and standard deviation of the five demos. Four demonstrations (D1, D2, D4, D5) achieved mean scores above 4.0, indicating that the majority of participants perceived the multimodal presentations as appropriate and effective. D3 showed the lowest average score (mean = 3.33), consistent with its broader distribution in Figure 23. Standard deviations fall between 0.5 and 0.8, suggesting moderate inter-participant variability.

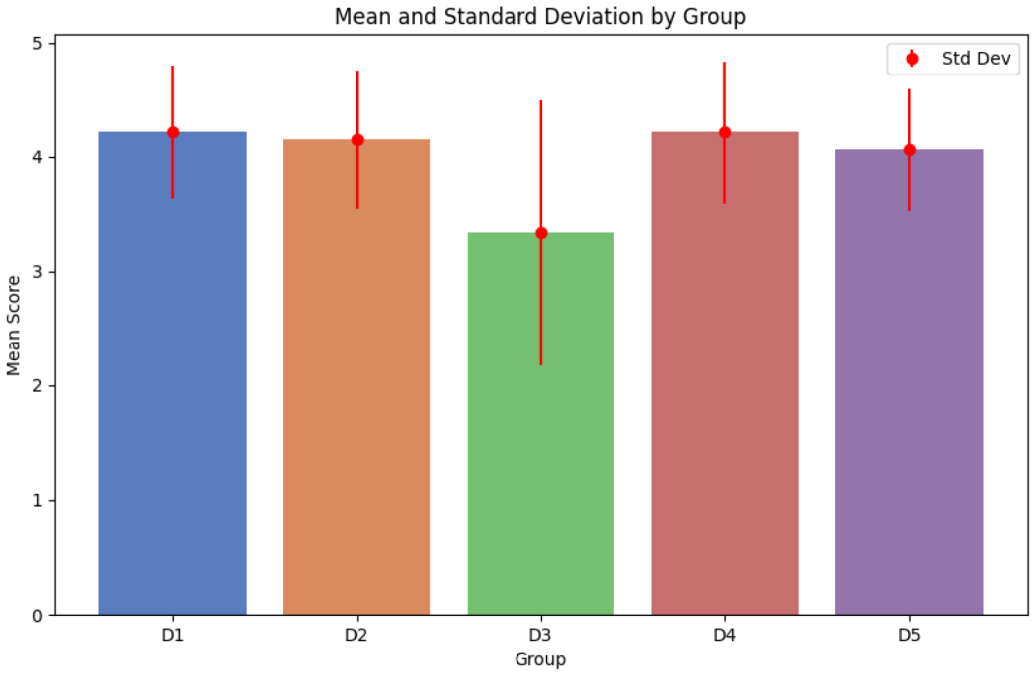



**Figure 24.** Mean and standard deviation of user ratings for the five demonstrations. Error bars indicate standard deviations.

#### 4.5.3. Summary

The results of Experiment E demonstrate that the proposed multimodal haiku presentation is generally effective and well-received. Participants reported that the combination of text, illustration, and background music helped them intuitively grasp the emotional tone and atmospheric qualities of the haiku. This combination offered a richer appreciation experience than reading the text alone. Four demonstrations (D1, D2, D4, and D5) achieved consistently high scores, further confirming the usefulness of the integrated haiku–image–music presentation.

In contrast, D3 received the lowest mean rating and exhibited the largest variance, indicating weaker alignment between the poem and its accompanying modalities. This pattern is consistent with the broader score distribution shown in Figure 23. Free-form comments also support this interpretation: several participants described the music in D3 as "slightly mismatched," and others felt that the illustration was "less representative of the poetic imagery." These observations suggest that certain poems, particularly those with subtle emotional shifts or less prototypical seasonal cues, may require more flexible or context-sensitive alignment mechanisms.

A likely source of this variability lies in the system's use of fixed weights for *Valence*, *Energy*, and *Dynamism* across all haiku. While this global weighting scheme works well for most poems, it may not fully capture the nuanced affective structures of more complex haiku. User feedback also indicates that culturally grounded cues (e.g., seasonal motifs, genre-specific expression patterns) play an important role in shaping perceived appropriateness.

Taken together, these findings point to promising directions for refinement, including adaptive weighting strategies, improved mapping between textual cues and musical features, and additional constraints in the image-generation pipeline. Experiment E thus complements the analytical validation of Experiments A to D by showing that the proposed cross-modal system is not only structurally valid but also capable of producing multimodal outputs that users find engaging, informative, and helpful for their haiku appreciation.

## 5. Discussion

The study investigated the potential for representing the aesthetic and emotional dimensions of haiku within a low-dimensional affective space derived from intelligent data analysis. Additionally, it examined whether such representations could facilitate coherent cross-modal alignment between text, music, and imagery. Across a series of analytical and user-centered evaluations (our validation framework), the results suggest that the proposed framework is both structurally sound and perceptually meaningful. At the same time, several limitations reveal significant challenges for modeling poetic content and designing multimodal appreciation systems for haiku appreciation.

### *5.1. Coherence of the Affective Space*

A key finding across the experiments is the emergence of a stable affective structure within the haiku corpus and the music dataset. Despite the brevity and metaphorical nature of haiku, the VED representation derived from our BERT-based intelligent data analysis (Experiment A) produced systematic patterns consistent with the traditional aesthetics of Japanese poetry. Likewise, the VET representation of instrumental music (Experiment C) displayed an interpretable geometry, with *Valence*, *Energy*, and *Tempo* forming meaningful discriminative axes.

The strong cross-modal correspondence observed in the axis-wise correlations (Experiment D-1) indicates that these textual and musical affective spaces are not independent. Instead, they appear to share a compatible emotional geometry. This alignment is noteworthy given that the two modalities were processed separately using different data analysis principles.

### *5.2. Cross-Modal Alignment and Multimodal Interpretation*

The nearest-neighbor matching between haiku and music produced high correlations across all affective dimensions (Experiment D-1) and non-random consistency at the cluster level (Experiment D-2). These results demonstrate that a simple geometric matching scheme, based solely on affective similarity, can generate a harmonious multimodal experience that users perceive as coherent and emotionally appropriate.

The user evaluation (Experiment E) further supports this conclusion. Most demonstrations were rated positively, and free-form comments indicate that participants found the cross-modal outputs helpful for understanding the emotional atmosphere of each poem. These findings suggest that affective similarity, even when derived from minimal feature sets, can serve as an effective mechanism for multimodal haiku appreciation.

### *5.3. Interpretability and Cultural Variability*

Although the overall reception was positive, variability in the ratings (notably for Demo 3 in Experiment E) indicates that affective alignment alone is not always sufficient. Haiku often rely on cultural cues, seasonal conventions, and highly compressed metaphorical associations. General-purpose language models have a challenge in capturing these elements, which can result in mismatches between generated images, selected music, and user expectations.

Cross-cultural interpretation also played a role. The participant group included both Japanese students and international students studying in Japan. Their differing cultural backgrounds occasionally produced divergent impressions of the same multimodal presentation. It suggests that affective alignment may require conditioning based on cultural or experiential factors. It highlights a broader challenge for the cross-cultural application of AI-based art interpretation: emotional resonance is partly universal but also significantly influenced by cultural norms and literary tradition.

### *5.4. Limitations and Future Challenges*

Several limitations of the present study offer opportunities for improvement.

#### 5.4.1. Limited Textual Domain

The haiku corpus used in this study consisted exclusively of the 1067 poems written by Matsuo Bashō. Although Bashō is foundational to the haiku tradition, his works occupy a relatively narrow emotional and stylistic space. To explore the feasibility of supervised emotion classification, we previously conducted a small annotation study in which several participants assigned one of four categorical labels (Joy, Anger, Sadness, Calmness) to each poem. Despite minor individual differences, a striking pattern emerged: the majority of annotators assigned the label "Calmness" to an overwhelming proportion of poems, often exceeding 70%. This implies that a trivial classifier labeling every haiku as "Calmness" would achieve a deceptively high baseline accuracy. It would render supervised categorical emotion prediction uninformative and unsuitable for this genre.

This bias is not unique to Bashō. Haiku as a literary form, and classical Japanese poetry more broadly, tends to gravitate toward subtle, contemplative, and low-arousal affective states. Such tendencies reflect deeper cultural aesthetics such as *mono no aware* (the pathos of transience) and *wabi–sabi* (the beauty of imperfection and austerity). Given these characteristics, categorical affect labels fail to capture the nuanced and predominantly tranquil emotional spectrum of haiku, which motivated our shift toward continuous, low-dimensional affective representations (our VED framework) rather than discrete emotion categories.

However, we acknowledge that relying on a single author introduces a potential stylistic bias. Consequently, the current system may be overfit to Bashō's specific aesthetic preferences. In this study, we intentionally position the system as a controlled proof of concept, focusing on the most representative figure of the genre. Therefore, at this stage, we do not claim generalization across different authors or cultures. Furthermore, this focus on tranquility results in a sparse distribution in the high-arousal regions of the VED space. Future work must incorporate diverse corpora to validate the broader applicability of the proposed representation. Specifically, including *waka* could introduce greater variance in emotional polarity due to its romantic themes. Meanwhile, classical Chinese poetry, which often depicts grandeur or social turbulence, would expand the dynamic range of *Energy*. Validating the VED framework on such diverse corpora is essential to ensure its cross-cultural generalizability beyond the specific aesthetic of Bashō.

#### 5.4.2. Dependence on Non-Curated Music Categories

The music dataset relied on Free BGM DOVA-SYNDROME's "Japanese-style" tag, which includes both traditional-sounding pieces and modern high-energy genres such as rock and metal. Although these styles expand the expressive range, they also introduce noise into the affective feature space. More granular genre control or expert-based music curation may further improve cross-modal matching quality.

#### 5.4.3. Fixed Affective Weighting

The current system uses a fixed weighting scheme for VED/VET similarity. The subjective feedback from Experiment E, especially the variability observed in Demo 3, suggests that users are sensitive to mismatches when poems contain subtle emotional shifts. To address this, future iterations could employ an adaptive weighting mechanism. For example, an attention-based module could dynamically increase the weight of *Dynamism* when high-motion verbs are detected, or prioritize *Valence* when sentiment analysis indi-

cates strong polarity. Such context-aware adjustment would likely reduce misalignment in complex or non-prototypical haiku.

#### 5.4.4. Absence of Expert Baselines

While all participants (Experiment E) were familiar with haiku from school education, none were professional poets or literary specialists. Their interpretations and free-form comments were informative but limited. Future work should compare them with expert judgments to strengthen the validation of literary and aesthetic alignment. Unlike general users who evaluate intuitive harmony, experts can assess the system using specialized metrics, such as "Seasonal Relevance," "Cultural Fidelity" to traditional aesthetics, and "Interpretive Depth." These metrics would help determine whether the system captures the profound subtext of the poem, such as *wabi-sabi*, rather than merely its surface-level semantics.

#### 5.4.5. Prompt-based Image Generation Limits

The generated watercolor images were generally well-received in Experiment E; however, achieving fine-grained control over composition, kigo depiction, and culturally specific motifs remains challenging. Current diffusion-based and instruction-following image models may struggle to capture traditional Japanese aesthetics without additional fine-tuning or style conditioning. Furthermore, as the system relies on large-scale transformer models, the outputs are subject to exposure bias and distributional shift, where the generated content may diverge from the intended distribution of traditional poetic imagery [30]. This highlights the need for more robust generation techniques that can better align with specific cultural domains.

To mitigate these issues, prompt engineering can be refined in several concrete ways. First, prompts may include explicit stylistic descriptors rooted in classical Japanese art, such as *wabi-sabi*, *yūgen*, traditional ink-wash painting, or *ukiyo-e* composition, to provide clearer cultural grounding. Second, the kigo component can be strengthened by linking each seasonal word to its canonical visual motifs (e.g., plum blossoms, harvest moon, first snow), reducing the model's reliance on implicit inference. Integrating a retrieval-augmented generation (RAG) approach or a lookup table grounded in a *Saijiki* (season-word dictionary) would be a practical extension to improve cultural authenticity. Finally, adopting a more structured prompt format with separate slots for emotional tone, kigo motifs, artistic style, and composition could improve reproducibility and aesthetic alignment across generated images.

### *5.5. Broader Implications and Cross-Cultural Extension Potential*

Despite these limitations, the findings demonstrate that affective representations, derived from intelligent data analysis, offer a promising basis for multimodal literary interpretation. The alignment observed in this study suggests that emotional structure can act as a cross-modal bridge between text, sound, and visual imagery, even in highly compact poetic forms such as haiku.

Therefore, this work contributes to the growing body of research that explores the application of intelligent data analysis to literary appreciation, the interpretation of multimodal art, and culturally grounded generative systems. Given this framework's potential for cross-cultural applications, future extensions may incorporate larger haiku corpora, other traditional Japanese poetic forms (such as waka and tanka), and adaptive systems that respond to user preferences and cultural backgrounds. Furthermore, broadening the scope beyond haiku may yield richer affective diversity. Classical Chinese poetry, for example, encompasses a much wider emotional range and a broader variety of themes, offering a valuable testbed for evaluating whether our affective representations generalize across literary traditions.

## 6. Conclusions

This study investigated whether the aesthetic and emotional qualities of haiku could be expressed through an intelligent data analysis-derived compact affective space and used to create a harmonious, multimodal experience involving text, music, and imagery. By integrating BERT-based emotional features (our intelligent data analysis component), acoustic affect descriptors, and controlled prompt-driven illustration, the proposed system demonstrates that affective similarity alone can serve as a viable application for enhancing cross-modal poetic presentation.

Through five complementary experiments, we validated the system from multiple perspectives. Experiment A demonstrated that the VED representation extracted from BERT-based contextual embeddings captures interpretable emotional structure within the Bashō haiku corpus. Experiment B confirmed that BERT provides superior stability and semantic fidelity compared with TF-IDF and Word2Vec as the basis for this analysis. Experiment C established the reliability of the musical VET descriptors, and Experiment D revealed strong cross-modal structural consistency between the textual and musical affective spaces. Finally, Experiment E demonstrated that the combined multimodal outputs were positively evaluated by users, indicating that affect-guided alignment enhances intuitive poetic appreciation.

These findings collectively suggest that low-dimensional affective spaces, when derived from intelligent data analysis, provide a practical, interpretable, and modality-independent foundation for applications in multimodal poetic appreciation. Although the current validation used Bashō's haiku as a controlled proof of concept, the proposed methodology is adaptable. It operates as a general, modality-independent framework that can coherently map text, sound, and visual imagery.

Several promising directions emerge from this work. To address the limitation of the single-author dataset and realize the full cross-cultural extension potential of this framework, future extensions may incorporate larger and more diverse poetic corpora, including waka, tanka, and classical Chinese poetry, to broaden the affective range and reduce genre-specific bias. More adaptive alignment mechanisms, capable of adjusting feature weights based on poem-specific cues or user preferences, may further improve cross-modal coherence. Advances in vision-language models also offer opportunities for more faithful and culturally sensitive illustration. Finally, expanding the validation to include expert poets, literary scholars, and diverse cross-cultural audiences will help establish a more comprehensive understanding of multimodal poetic appreciation.

Overall, this work provides methodological insights and empirical evidence indicating that affect-guided multimodal generation could be a valuable approach for intelligent data analysis applications in poetic appreciation. The developed framework is an initial step toward systems integrating emotion, culture, and creativity across modalities to support more accessible and engaging poetic art experiences.

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

**Funding:** This research was partially supported by JSPS KAKENHI Grant Numbers JP21K17862 and JP25K15354.

**Institutional Review Board Statement:** Ethical review and approval were waived for this study due to its nature as an anonymous user questionnaire survey, which does not involve personal data collection or invasive procedures, in accordance with the "Ethical Guidelines for Medical and Biological Research Involving Human Subjects" established by the Japanese Ministry of Health, Labour, and Welfare (MHLW) and the "Regulations for General Research Involving Human Subjects" at Yamaguchi University.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The datasets used in this study are openly available as follows: The bashoDB (Bashō Haiku Database) provided by Yamanashi Prefectural University, available at http://www2.yamanashi-ken.ac.jp/~itoyo/basho/basho.htm (accessed on 18 November 2025); and the Free BGM DOVA-SYNDROME music library, available at https://dova-s.jp/ (accessed on 18 November 2025). Further inquiries can be directed to the corresponding author.

**Acknowledgments:** The authors gratefully acknowledge the providers of the publicly available datasets and tools used in this study. These include the bashoDB (Bashō Haiku Database) from the University of Yamanashi, seasonal term dictionaries from Weblio, BERT models from Tohoku University, and music from Free BGM DOVA-SYNDROME.

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

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| AI | Artificial intelligence |
| ARI | Adjusted Rand Index |
| BPM | Beats-per-minute |
| LLMs | Large language models |
| MER | Music emotion recognition |
| MFCCs | Mel-frequency cepstral coefficients |
| NLP | Natural language processing |
| PCA | Principal component analysis |
| RAG | Retrieval-augmented generation |
| RMS | Root-mean-square |
| STFT | Short-time fourier transform |
| ZCR | Zero-crossing rate |

## Appendix A. Haiku Used in the Image-Reproducibility Experiment

Table A1 lists the six representative haiku used in Experiment D-3 (Section 4.4.3). These haiku were chosen to cover all six seasonal categories.

**Table A1.** Haiku metadata used for constructing prompts in Experiment D-3.

| ID | Original Text | English Translation | Season Name | Season-Specific Scene | Season-Specific Color | Cluster-Specific Emotional Tone |
|---|---|---|---|---|---|---|
| 839 | 古池や蛙飛びこむ水の音 | Old pond—a frog leaps in, the sound of water. | Spring | A tranquil spring landscape with cherry blossoms and fresh greenery. | Soft pastel pink and light green. | High-energy, tense mood. |
| 648 | 夏草や兵どもが夢の跡 | Summer grasses—all that remains of warriors' dreams. | Summer | A bright summer scene near water, with deep blue sky and strong sunlight. | Vivid blue and warm yellow. | Low-energy, calm mood. |

**Table A1.** *Cont.*

| ID | Original Text | English Translation | Season Name | Season-Specific Scene | Season-Specific Color | Cluster-Specific Emotional Tone |
|---|---|---|---|---|---|---|
| 932 | 名月や池をめぐりて夜もすがら | Bright autumn moon—circling the pond all through the night. | Autumn | An autumn landscape with maple leaves and a slightly nostalgic atmosphere. | Deep orange, red and brown. | Low-energy, calm mood. |
| 78 | いざさらば雪見にころぶ所まで | Come, let us go—until we tumble in the snow while viewing it. | Winter | A quiet winter scene with snow and a cold, clear atmosphere. | White, pale blue and muted gray. | High-energy, tense mood. |
| 231 | 徒歩ならば杖突坂を落馬かな | Had I been on foot, I would not have fallen from my horse on Tsue-tsuki Slope. | All Year | A simple natural landscape that could belong to any season. | Balanced natural colors. | Low-energy, calm mood. |
| 234 | 門松やおもへば一夜三十年 | New Year's pine—thinking back, one night feels like thirty years. | New Year | A calm New Year scene with subtle festive atmosphere. | Soft red and white. | High-energy, tense mood. |

## References


1. Rowland, P. New Directions in English-language Haiku: An Overview and Assessment. *Iafor J. Lit. Librariansh.* **2013**, *2*, 53–66. [CrossRef]
2. Kawamura, H.; Yamashita, M.; Yokoyama, S. *Artificial Intelligence Reads Haiku: The Challenge of AI Issa-kun*; Ohmsha, Ltd.: Tokyo, Japan, 2021.
3. Yokoyama, S.; Yamashita, T.; Kawamura, H. Generation and Selection of Haiku Poems Using Deep Learning. *J. Jpn. Soc. Artif. Intell.* **2019**, *34*, 467–474. [CrossRef]
4. Yokoyama, S.; Yamashita, T.; Kawamura, H. Development of a Selection Mechanism Using Deep Auto-Regression Model with Examples of Seasonal Fixed-form Haiku in Japanese. In Proceedings of the 38th Annual Conference of the Japanese Society for Artificial Intelligence, Online, 28–31 May 2024. [CrossRef]
5. Yuki, T.; Ueda, N.; Oka, T.; Komachi, M. Toward Automatic Evaluation of Haiku Using GPT. In Proceedings of the 38th Annual Conference of the Japanese Society for Artificial Intelligence, Online, 28–31 May 2024. [CrossRef]
6. Hanano, A.; Yokoyama, S.; Yamashita, T.; Kawamura, H. Evaluation of Haiku using Masked Language Model RoBERTa. In Proceedings of the 84th National Convention of the Information Processing Society of Japan, Online, 3–5 March 2022; Volume 2022; pp. 1061–1062.
7. Tateishi, K.; Mizuguchi, S. Are we biased against AI-made haiku poems? *Proc. Linguist. Soc. Am.* **2025**, *10*, 1–15. [CrossRef]
8. Russell, J.A. A Circumplex Model of Affect. *J. Personal. Soc. Psychol.* **1980**, *39*, 1161–1178. [CrossRef]
9. Shahriar, S.; Al Roken, N.; Zualkernan, I. Classification of Arabic Poetry Emotions Using Deep Learning. *Computers* **2023**, *12*, 89. [CrossRef]
10. Walunj, V.; Choudhary, A.; Kale, A.; Gade, M. Emotion Recognition from Formal Text (Poetry). *Int. J. Adv. Res. Comput. Commun. Eng.* **2023**, *12*, 1469–1473. [CrossRef]
11. Ohmameuda, T. Preliminary Study on Automatic Extraction of Seasonal Words (Kigo) in Haiku. *Gunma-Kosen Rev.* **2023**, *42*, 95–100. [CrossRef]
12. Kim, Y.; Schmidt, E.; Migneco, R.; Morton, B.; Richardson, P.; Scott, J.; Speck, J.; Turnbull, D. Music emotion recognition: A state of the art review. In Proceedings of the 11th International Society for Music Information Retrieval Conference, ISMIR, Utrecht, The Netherlands, 9–13 August 2010; pp. 255–266.
13. Yang, Y.H.; Chen, H.H. *Music Emotion Recognition*; CRC Press: Boca Raton, FL, USA, 2012. [CrossRef]
14. Zhao, J.; Yoshii, K. Multimodal Multifaceted Music Emotion Recognition Based on Self-Attentive Fusion of Psychology-Inspired Symbolic and Acoustic Features. In Proceedings of the 2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Taipei, Taiwan, 31 October–3 November 2023; pp. 1641–1645. [CrossRef]
15. Cui, X.; Wu, Y.; Wu, J.; You, Z.; Xiahou, J.; Ouyang, M. A review: Music-emotion recognition and analysis based on EEG signals. *Front. Neuroinformatics* **2022**, *16*, 997282. [CrossRef] [PubMed]
16. Zhao, Y.; Yang, M.; Lin, Y.; Zhang, X.; Shi, F.; Wang, Z.; Ding, J.; Ning, H. AI-Enabled Text-to-Music Generation: A Comprehensive Review of Methods, Frameworks, and Future Directions. *Electronics* **2025**, *14*, 1197. [CrossRef]
17. Mao, Z.; Zhao, M.; Wu, Q.; Zhong, Z.; Liao, W.H.; Wakaki, H.; Mitsufuji, Y. Cross-Modal Learning for Music-to-Music-Video Description Generation. *arXiv* **2025**, arXiv:2503.11190.

18. Minato, K.; Onai, R. A prototype system of composite image generation for Haiku. *Ite Tech. Rep.* **2009**, *33*, 43–46. (In Japanese) [CrossRef]
19. Minato, K.; Okabe, M.; Onai, R. A Prototype System for Generating Composite Images Suitable for Haiku. In Proceedings of the Workshop on Interactive Systems and Software (WISS), Hong Kong, China, 12–14 December 2010; Information Processing Society of Japan (IPSJ): Tokyo, Japan, 2010; p. D37.
20. Toda, K.; Ito, A.; Yoshino, T. Investigation of an Automatic Generation System of Images Adapted to Haiku Scenes. In Proceedings of the 2023 IPSJ Kansai Chapter Annual Convention, Osaka, Japan, 19 August 2023; Volume 2023, 7p.
21. Ramesh, A.; Pavlov, M.; Goh, G.; Gray, S.; Voss, C.; Radford, A.; Chen, M.; Sutskever, I. Zero-Shot Text-to-Image Generation. *arXiv* **2021**, arXiv:2102.12092.
22. Sasaki, S.; Ushiama, T. A Method of Generating Images to Represent Impressions of a Piece of Music Using the Text2Image Model. In Proceedings of the IEICE Technical Committee on Data Engineering, Kitakyushu International Conference Center, Sapporo, Japan, 29 November–1 December 2023; IEICE Technical Report: DE2023-26 (No. DE-192); Volume 123, pp. 89–94.
23. Liu, B.; Fu, J.; Kato, M.P.; Yoshikawa, M. Beyond Narrative Description: Generating Poetry from Images by Multi-Adversarial Training. In Proceedings of the ACM International Conference on Multimedia, Seoul, Republic of Korea, 22–26 October 2018; ACM: New York, NY, USA, 2018; pp. 783–791. [CrossRef]
24. Tosa, N. *Cross-Cultural Computing: An Artist's Journey*; Springer Series on Cultural Computing; Springer: London, UK, 2016. [CrossRef]
25. Ohno, M.; Yokosawa, K.; Narumi, T. Basics of Integrated Perception and its Application to the Multi-Sensory Interface. *Trans. Virtual Real. Soc. Jpn.* **2022**, *27*, 18–28. [CrossRef]
26. Lu, M.Y.; Chen, B.; Williamson, D.F.K.; Chen, R.J.; Zhao, M.; Chow, A.K.; Ikemura, K.; Kim, A.; Pouli, D.; Patel, A.; et al. A Multimodal Generative AI Copilot for Human Pathology. *Nature* **2024**, *634*, 466–473. [CrossRef] [PubMed]
27. Reimers, N.; Gurevych, I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; Inui, K., Jiang, J., Ng, V., Wan, X., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 3982–3992. [CrossRef]
28. Ethayarajh, K. How Contextual Are Contextualized Word Representations? Comparing the Geometry of BERT, ELMo, and GPT-2 Embeddings. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; Inui, K., Jiang, J., Ng, V., Wan, X., Eds.; pp. 55–65. [CrossRef]
29. Eerola, T.; Vuoskoski, J.K. A comparison of the discrete and dimensional models of emotion in music. *Psychol. Music* **2011**, *39*, 18–49. [CrossRef]
30. Pozzi, A.; Incremona, A.; Tessera, D.; Toti, D. Mitigating exposure bias in large language model distillation: An imitation learning approach. *Neural Comput. Appl.* **2025**, *37*, 12013–12029. [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.

electronics

MDPI

*Article*

# Dynamic Topic Analysis and Visual Analytics for Trajectory Data: A Spatial Embedding Approach

**Huarong Chen [1,2], Yadong Wu [3,*], Jing Lei [4], Weixin Zhao [5], Guijuan Wang [2], Jing Liao [2] and Fupan Wang [2]**

[1] School of Information and Control Engineering, Southwest University of Science and Technology, Mianyang 621010, China
[2] School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang 621010, China
[3] School of Computer Science and Engineering, Sichuan University of Science and Engineering, Yibin 644007, China
[4] Technical Center, Mianyang Xinchen Engine Co., Ltd., Mianyang 621000, China
[5] College of Computer Science, Sichuan University, Chengdu 610065, China
* Correspondence: wuyadong@suse.edu.cn

**Abstract**

Analyzing the evolution of trajectory topics is fundamental to understanding urban mobility and human activity. Existing methods, however, often struggle to capture complex spatio-temporal semantics and are constrained by fixed time windows, which limits multi-scale temporal analysis. This paper presents a novel method to model the dynamic topics of trajectories to address these limitations. The proposed method combines a domain-specific trajectory embedding strategy, a flexible dynamic topic modeling pipeline, and an interactive visualization system to address these limitations. Firstly, the method introduces a novel embedding method that uses a retrained RoBERTa model with a word-level tokenizer on Morton-coded trajectories to effectively learn spatial context and sequential patterns. Secondly, a BERTopic-based approach is employed for topic modeling, featuring an adjustable time window that allows for flexible analysis of topic dynamics across different temporal scales without model retraining. Furthermore, an interactive visualization system with coordinated spatio-temporal views translates abstract model outputs into an intuitive format, enabling direct exploration of evolving trajectory topics. Experiments on a large-scale taxi trajectory dataset demonstrate the proposed method's effectiveness in identifying coherent and meaningful patterns of dynamic trajectory topics.

**Keywords:** dynamic topic modeling; spatio-temporal data mining; trajectory data; visual analysis

## 1. Introduction

Trajectory data hold rich information about spatio-temporal behavior [1]. Analyzing trajectory data is fundamental to understanding urban mobility and human activity [2]. The underlying patterns, or topics, within trajectory data are not static. Trajectory topics often correlate with temporal cycles. For instance, commute flows differ significantly between weekdays and non-workdays. An understanding of dynamic changes is essential to applications in traffic management, public service optimization, and urban planning. A clear need, therefore, exists for methods to effectively model and interpret trajectory topic evolution.

Modeling trajectories to effectively capture features is a fundamental step for analyzing dynamic topics. Existing modeling approaches often use representation learning

to generate high-dimensional embeddings for trajectory sequences. RNN-based methods [3] employ a recurrent structure to process sequences. However, for long trajectories, the recurrent structure struggles to propagate information from the start to the end of a sequence [4–7]. Transformer-based methods [8] offer an alternative but require supervised training for specific tasks. Consequently, the resulting trajectory representations become task-dependent [9–13]. A new trajectory embedding strategy is required to produce effective semantic representations that preserve spatial context.

Many existing topic modeling methods lack flexibility for analyzing topic dynamics across different temporal scales. These methods often rely on fixed time windows for analysis. An analysis based on fixed windows limits the scope of inquiry. A predefined interval may be too coarse to capture short-term events or too fine to reveal long-term trends. Such limitations prevent a comprehensive, multi-scale understanding of movement dynamics. The dynamic topic model (DTM) [14], a dynamic extension of Latent Dirichlet Allocation (LDA) [15], exemplifies these challenges. While DTM allows topics to evolve over time, the model requires a fixed number of time slices to be set before analysis. This requirement reduces flexibility, especially for temporally uneven data. Any adjustment to the modeling goal in DTM requires a full retraining of the model with a new time slice configuration. Such retraining is computationally expensive. DTM also requires pre-specifying the number of topics and offers limited semantic modeling capabilities. A clear need, therefore, exists for a more flexible topic modeling approach. Such an approach should permit dynamic adjustment of the temporal scope to match different behavioral patterns and analytical goals.

Visual analytics systems use interactive interfaces to help users explore and interpret dynamic trajectory topics [16]. Existing visualization tools for trajectory analysis often focus on presenting statistical information [17–19]. A key limitation is the lack of support for exploring the various components of a topic to understand its underlying meaning. Effective visual analysis of trajectory topics thus requires intuitive components designed for interactively exploring multi-level semantics.

To address these challenges, this paper presents a new method for constructing dynamic topic models from trajectory data. The proposed method consists of three main components. First, the method introduces a new embedding strategy to capture complex trajectory semantics while preserving spatial context. This strategy uses a modified RoBERTa model and a word-level tokenizer to generate context-aware embeddings from Morton-coded trajectories. Second, the method employs BERTopic [20], a data-driven neural topic model, for a more flexible modeling of dynamic topics. BERTopic clusters the trajectory embeddings and then generates topic representations using a class-based TF-IDF procedure. Then, a visualization system centered on trajectory topic words and features is developed for the interactive exploration of dynamic movement topics. The system translates abstract model outputs into an intuitive interface, enabling the direct exploration and interpretation of evolving trajectory topics. The proposed method applies established text processing methods to the challenge of topic modeling for unstructured spatio-temporal data. The main contributions of this work are as follows:

(1) A domain-specific trajectory embedding method to effectively capture both spatial context and sequential patterns. This addresses the limitations of information decay in RNN-based approaches and task-dependent representations in standard Transformer models.

(2) A flexible dynamic topic modeling method using an adjustable time window to support multi-scale temporal analysis. This overcomes the constraint of fixed time slices found in traditional dynamic topic models.

(3) An interactive visualization system to translate abstract model outputs into an intuitive interface for the direct exploration and interpretation of evolving trajectory topics.

The rest of the paper is organized as follows: We introduce related work on modeling dynamic topics of trajectories in Section 2. Then, we give the detail of the proposed method for dynamic modeling trajectory topics in Section 3. Extensive experiments are performed in Section 4. Next, we present a discussion and mention future work in Section 5. Finally, we conclude this paper in Section 6.

## 2. Related Work

In this section, we briefly introduce related works. Specifically, we introduce works on the dynamic topics of trajectories first, and then, we introduce works of modeling trajectories. Finally, we detail related work about trajectory visualization.

Methods for modeling dynamic trajectory topics are generally classified into static and dynamic approaches. Static topic models such as LDA [15], NMF [21], and BTM [22] can identify global movement patterns across an entire dataset [23–28]. A primary limitation of static models is the inability to capture temporal topic dynamics. Dynamic extensions like Dynamic NMF [29] and dynamic topic models (DTMs) [14] were developed to explicitly model topic evolution. However, a common limitation is the requirement to fix the number and duration of time slices prior to model training. Furthermore, traditional topic models including LDA and DTMs employ a bag-of-words assumption. This assumption ignores the sequential order of trajectory points, preventing the capture of complex spatial movement semantics. BERTopic [20] presents a modern deep learning paradigm for this problem. The method adapts text modeling techniques to capture both sequential patterns and spatial context in trajectory data. Moreover, BERTopic enhances analytical flexibility through a modeling process that supports adjustable time windows. Unlike traditional DTMs, which require computationally expensive retraining whenever time slices are altered, this approach allows for dynamic window adjustments during the inference phase without retraining, thereby offering superior computational efficiency for multi-scale analysis. Representation learning is a key component within BERTopic for capturing sequence features. The following section details relevant methods for trajectory sequence modeling.

Effective trajectory modeling forms the basis for dynamic topic analysis. Existing methods for learning trajectory representations [30] are generally classified into two categories: RNN-based [3] and Transformer-based [8] approaches. RNN-based methods [4–7] use a recurrent structure to process sequential information. For long trajectory sequences, however, this structure struggles to propagate information from the start to the end of a sequence, resulting in information decay. Standard Transformer-based methods [9–13] typically require supervised training on downstream tasks. The resulting trajectory representations are, therefore, task-specific. Consequently, applying these methods to trajectory topic modeling introduces the risks of information decay and task-dependent feature representations. To address these limitations, the proposed method leverages RoBERTa [31]. The approach combines RoBERTa with structural encoding and a word-level tokenizer [32] to learn contextual trajectory embeddings while preserving the semantic independence of individual locations.

Converting abstract trajectory topics into interpretable insights depends on effective visualization. Visual analytics is a common approach for exploring features from trajectory data [17–19]. Many existing methods use visual components to analyze traffic trajectories. For instance, He et al. [33] developed a visual filter, a form of coordination between views, to explore the spatio-temporal evolution of topics. Gao et al. [34] employed visualization to guide the trajectory clustering process. However, a comprehensive interpretation of dynamic topics requires integrating a topic's multi-level semantics, from individual topic words and trajectory segments to the overall theme. Existing approaches often lack this deep integration needed to explain dynamic topic behavior. The proposed visualization

system is designed based on the core attributes of trajectories. The approach presents trajectory topics by centering on two semantic levels: topic words and trajectory segments. A system of coordinated views displays topic features across different time slices. This design yields clear visual representations of topic characteristics and allows analysts to explore dynamic topic evolution from macro- to micro-scales.

## 3. Methodology

The primary research problem is to capture dynamic semantics from unstructured trajectory data. The main goal is to build a flexible framework for analyzing topic evolution. Based on this goal, we introduce the details of our dynamic topic modeling method for trajectory data in this section. Figure 1 illustrates the pipeline of the proposed method. There are three steps to modeling dynamic topics in trajectory data and analyzing the dynamic topics in trajectory data with visualization. Basically, after generating trajectory embeddings with a domain-specific embedding strategy, we incorporate the adjustable topic modeling method to generate trajectory dynamic topics. Then, we develop a visualization system to interactively analyze the details of the dynamic topics of the trajectory.

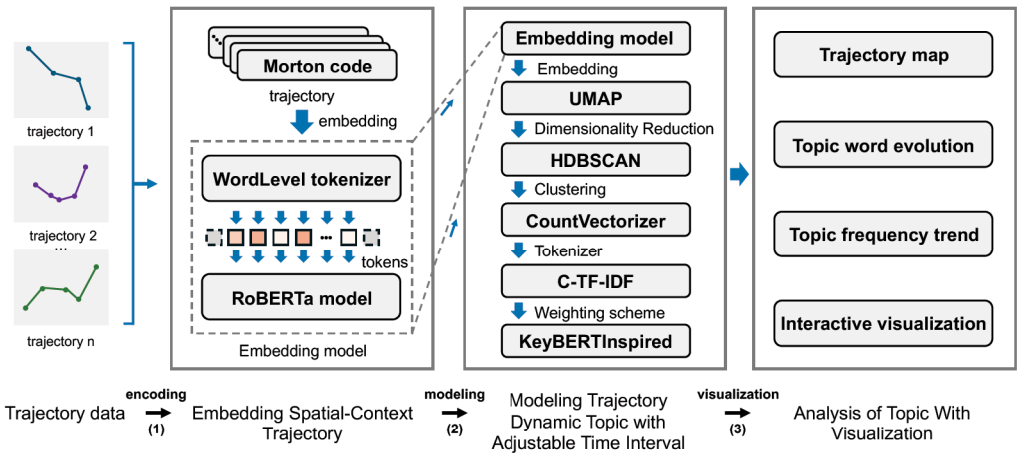



**Figure 1.** The pipeline of proposed dynamic topic modeling method for trajectory data: (1) trajectory embedding with spatial-context feature, (2) dynamic trajectory topic modeling with adjustable time interval, and (3) visualization for dynamic topic modeling.

### *3.1. Trajectory Embedding with Spatial-Context Feature*

#### 3.1.1. Morton Encoding for Spatial Discretization

The trajectory is composed of a series of trajectory coordinates in sequential order, and it is necessary to simplify the trajectory data structure. The trajectory coordinates are encoded by Morton codes [35]. The following formulation is the definition of the Z-order curve of Morton codes:

$$\text{Morton}(x, y) = \sum_{i=0}^{l-1} \left( x_i 2^{2i} + y_i 2^{2i+1} \right), \tag{1}$$

where $x$ and $y$ are the longitude and latitude coordinates in binary format, $x_i$ and $y_i$ represent the bit at position $i$, and $l$ is used to quantify the number of digits of the coordinates. Morton encoding simplifies the structure of trajectory data because its dimensions are encoded from a multi-dimensional space to a single-dimensional space. Basically, Morton encoding adapts the Z-curve to fill the space, which keeps the spatial information. The length of a Morton code remains 12 bits with an accuracy of 10 m. Additionally, Morton encoding provides the benefit of keeping the related spatial feature in nearby GPS

coordinates. In detail, near trajectory GPS coordinates are encoded into the near digital space, which improves the calculation efficiency. We convert all trajectory coordinates with the expression of longitude and latitude into the string word with Morton encoding, establishing the bridge from trajectory sequence to the following RoBERTa method.

#### 3.1.2. Tokenizer and Vocabulary Design

The tokenizer converts the trajectory sequence into a discrete signal sequence. The WordLevel tokenizer is suitable in the task of trajectory embedding because the WordLevel tokenizer preserves the integrity of Morton identifiers and domain terms. Additionally, The WordLevel tokenizer avoids the semantic damage caused by sub-word splitting and preserves the spatial encoding granularity.

The training process of WordLevel tokenizer is present in Algorithm 1. Let the trajectory be represented as a sequence $T = (g_1, g_2, ..., g_m)$, where each $g_i \in \mathcal{G}$ denotes a Morton code. Algorithm 1 aims to construct a WordLevel tokenizer $\Phi : \mathcal{T} \to \mathbb{Z}^d$ that satisfies

$$|\Phi(T)| = L_{\max}, \forall T \in \mathcal{T} \tag{2}$$

where the maximum sequence length is represented as $L_{\max}$. The training pipeline includes three phases. Firstly, we construct the vocabulary $\mathcal{V}_{\text{base}}$ by combining the special tokens $\mathcal{V}_{\text{spec}}$ with the unique Morton codes tokens $\mathcal{V}_{\text{geo}}$ extracted from all trajectories. Then, with the configurations of whitespace splitting at Morton code boundaries and the length limitation $L_{\max}$, the mapping function $\phi$ is constructed to initialize the WordLevel tokenizer $\Phi$.

**Algorithm 1** WordLevel tokenizer training.

**Require:** Trajectory set $\mathcal{T} = \{T_1, ..., T_n\}$, Max trajectory length $L_{\max}$, Vocab size $V_{\max} \geq 5$
**Ensure:** Tokenizer $\Phi$

$\mathcal{V}_{\text{geo}} \leftarrow \emptyset$
**for** each $T_i \in \mathcal{T}$ **do**
    $\mathcal{V}_{\text{geo}} \leftarrow \mathcal{V}_{\text{geo}} \cup \{g | g \in \text{split}(T_i)\}$
**end for**
$\mathcal{V}_{\text{spec}} \leftarrow \{\langle \text{s} \rangle, \langle \text{pad} \rangle, \langle \text{/s} \rangle, [\text{UNK}], [\text{MASK}]\}$ ▷ Fixed IDs 0–4
$\mathcal{V}_{\text{base}} \leftarrow \mathcal{V}_{\text{spec}} \cup \text{sort}(\mathcal{V}_{\text{geo}})$ ▷ Preserve ordering
Build vocabulary mapping: $\phi : \mathcal{V}_{\text{base}} \to \{0, \cdots, |\mathcal{V}_{\text{base}}| - 1\}$
    where $\phi(t_k) = k$ for $t_k \in \mathcal{V}_{\text{base}}$
Initialize $\Phi$ with:

- Tokenization function: $\Phi_{\text{tokenize}}(T) = \langle \text{s} \rangle \oplus \phi(g_1, ..., g_m) \oplus \langle \text{/s} \rangle$
- Preprocessing: Whitespace splitting
- Truncation/Padding: $\forall T, |\Phi(T)| = L_{\max}$

return $\Phi$

The WordLevel tokenizer has better adaptability to the trajectory. Compared with the BPE tokenizer, the WordLevel tokenizer generates the complete token from the trajectory sequence. The complete token of Morton codes maintains the local spatial pattern. The object with minimal byte character given by the BPE tokenizer causes difficulties in learning the trajectory pattern. In the experimental section, we compare the modeling performance of two topic modeling approaches. These two topic modeling approaches are respectively equipped with a BPE tokenizer and a word-level tokenizer.

#### 3.1.3. Domain-Specific RoBERTa Training

RoBERTa is a Transformer-based language model and presents an optimized iteration of BERT. RoBERTa benefits from the self-attention mechanism and generates a contextualized representation of words. To capture the spatial context of a trajectory, we train RoBERTa with the WordLevel tokenizer. We use a trajectory dataset to train RoBERTa and aggregate the outputs from RoBERTa by mean pooling to generate trajectory embeddings. Unlike general-purpose models, this retraining strategy allows the model to learn domain-specific features. The retrained RoBERTa model adapts the vocabulary distribution of trajectory text.

The RoBERTa model is trained with the Masked Language Modeling (MLM) strategy. When training the RoBERTa model, we randomly mask a trajectory coordinate with a special token [MASK] and make the RoBERTa model predict the masked trajectory coordinate. The MLM training strategy makes the RoBERTa model learn the sequence pattern and does not need any notation. Additionally, the task of mask prediction forces the model explore the context of trajectory and strengthens the understanding of the trajectory sequence. In addition, the random mask acts as the trajectory noise during the training process with the MLM strategy, increasing the robustness of model encoding. For a trajectory token $v_i$, the MLM loss is defined as

$$\mathcal{L}_{\text{MLM}} = -\frac{1}{|M|} \sum_{i \in M} \log P(v_i | \mathbf{h}_i^L), \tag{3}$$

where $M$ is the set of masked token indices, $v_i$ is the correct trajectory token, and $\mathbf{h}_i^L$ denotes the output vector corresponding to the masked position.

RoBERTa captures the spatial relationship between nearby trajectory coordinates and learns the semantics from the whole trajectory, which offers trajectory embeddings with rich semantic information. Additionally, the position encoding of RoBERTa can capture the order of trajectory coordinates, complementing the spatial proximity of Morton codes. In the Experiments Section, the topic modeling performance of two kinds of RoBERTa models, a RoBERTa model trained with trajectory data and a RoBERTa model fine-tuned with trajectory data, is compared.

### *3.2. Adjustable Dynamical Topic Modeling for Trajectory Data*

To overcome the limitations of the bag-of-words model and fixed time windows, we construct a flexible method for the dynamic topic modeling of trajectories based on BERTopic. During the inference phase, the capability for dynamic time window adjustment expands the boundaries of trajectory topic modeling. The core phases of the adjustable dynamic topic modeling method include embedding, dimensionality reduction, clustering, and topic representation. Let $\mathcal{D} = \{d_1, \cdots, d_N\}$ represent the document collection of a trajectory. Document $d$ could be split into the set of trajectory sequences $\mathcal{T}_d = (T_{d,1}, T_{d,2}, \cdots, T_{d,|\mathcal{T}_d|})$, where $|\mathcal{T}_d|$ is the sentence count of document $d$. We detail the adjustable dynamic topic modeling method in the following content.

#### 3.2.1. UMAP Dimensionality Reduction

Operating directly on raw trajectory embeddings from RoBERTa for modeling topics encounters dimensionality-induced obstacles. Basically, the high-dimensional embedding space characterizing trajectory data fundamentally hinders direct topic modeling by introducing semantic sparsity and computational intractability. UMAP achieves non-linear dimensionality reduction based on manifold learning, enabling tasks for dimension reduction for trajectory embeddings. We could divide UMAP into two main steps. Basically,

UMAP learns the manifold structure in high-dimensional data spaces and constructs its low-dimensional representation. As a result, the inherent manifold learning preserves the local and global patterns of the trajectory embeddings.

#### 3.2.2. HDBSCAN Density-Based Clustering

HDBSCAN performs hierarchical clustering on trajectory embeddings, yielding clusters of similar trajectory embeddings [36]. Let $\mathbf{z}_i$ denote a trajectory embedding generated from UMAP reduction. HDBSCAN processes trajectory embedding $\mathbf{z}_i$ and preserves the principal topological structures within the embedding space. HDBSCAN generates the cluster labels $t_i \in \{-1, 0, 1, \cdots, K\}$, where $-1$ denotes noise. HDBSCAN automatically determines the optimal number of clusters while avoiding parameter sensitivity issues. This reveals the core advantage of using HDBSCAN for trajectory topic clustering: it establishes membership relationships between trajectory-embedded documents and their corresponding clusters. Additionally, HDBSCAN detects semantically coherent trajectory topics by identifying core points to locate high-density regions and marking low-density areas as noise. HDBSCAN identifies arbitrarily shaped trajectory embeddings using density-reachable principles, effectively adapting to dynamically changing movement patterns.

#### 3.2.3. CountVectorizer Frequency Quantization

CountVectorizer is an essential step in trajectory topic calculation. It links the trajectory clustering results with keywords. Given the trajectory document $\mathcal{D}$ and vocabulary $\mathcal{V}_{\text{base}}$. CountVectorizer constructs a sparse matrix $X = [x_{t,i}] \in \mathbb{N}^{N \times V}$, where $x_{t,i}$ denotes the frequency of word $w_i$ in document $d_t$.

#### 3.2.4. c-TF-IDF Topic Representation

c-TF-IDF calculates the weight of trajectory words at the cluster level. c-TF-IDF takes as input the clusters $\{\mathcal{C}_1, \ldots, \mathcal{C}_{|T|}\}$ from the output of HDBSCAN, where each cluster $\mathcal{C}_t$ is a document collection. Then, term frequency $\text{tf}_{t,i}$ and inverse document frequency $\text{idf}_i$ for the trajectory words are computed, yielding the importance score $\omega_{t,i}$ for word $\omega_i$ in the cluster:

$$\omega_{t,i} = \text{tf}_{t,i} \cdot \text{idf}_i. \tag{4}$$

c-TF-IDF achieves semantic consistency and enhances discriminability through $\text{idf}_i$. For trajectory embeddings, the method applies normalization to eliminate the effect of length variation, enabling fair cross-topic comparisons. Additionally, c-TF-IDF employs $\text{idf}_i$ to suppress high-frequency geographical noise.

#### 3.2.5. KeyBERTInspired Contextual Topic Representation

KeyBERTInspired operates as an essential plug-in and a re-ranker for c-TF-IDF results to enhance the consistency of trajectory semantics. It ensures that the selected topic words closely represent the cluster centroid. Let $\mathcal{D}_t = \{d_{t,1}, \ldots, d_{t,n_t}\}$ denote the representation documents of the $t$-th topic and $\mathcal{K}_t = \{w_{t,1}, \ldots, w_{t,k_t}\}$ be candidate keywords, where $w_{t,k_t}$ is the $k_t$-th keyword in the $t$-th topic. KeyBERTInspired conducts the semantic ranking of candidate keywords $\mathcal{K}$ by computing spatial similarity between representative documents $\mathcal{D}$ and keyword embeddings. Let $\mathbf{e}(\cdot)$ denote the encoder from RoBERTa; the centroid of

the trajectory topic is calculated with $\mathbf{v}_t = \frac{1}{n_t}\sum_{i=1}^{n_t}\mathbf{e}(d_{t,i}) \in \mathbb{R}^d$. Subsequently, compute the cosine similarity between trajectory embeddings and centroids with

$$s_{t,j} = \frac{\mathbf{e}(w_{t,j})^\top \mathbf{v}_t}{\|\mathbf{e}(w_{t,j})\|_2 \|\mathbf{v}_t\|_2}, \tag{5}$$

where $j = 1, \ldots, k_t$. KeyBERTInspired selects keywords by maximizing the similarity scores $s_{t,j}$, enforcing directional alignment between trajectory words and the trajectory topic. The collaboration between KeyBERTInspired and c-TF-IDF improves trajectory topic modeling while maintaining computational efficiency and semantic precision.

#### 3.2.6. Dynamic Time-Window Parameterization

Our topic modeling method outperforms traditional approaches by dynamically adjusting the temporal range for analysis, optimizing time windows for different behavioral patterns. Time window $W_t$ is defined as

$$W_t = \{d_i | t_s \leq \text{timestamp}(d_i) < t_s + \Delta T\}, \tag{6}$$

where $d_k$ represents a trajectory document, $\text{timestamp}(d_i)$ is its associated time, $\Delta T$ denotes the size of the time window, and $t_s$ is the start time of the window. This formula defines a time window $W_t$ containing all events points $d_i$ meeting specific temporal conditions. Specifically, each $\text{timestamp}(d_i)$ must be included in $[t_s, t_s + \Delta T]$, where $t_s$ denotes the start time and $\Delta T$ the interval duration. Adjustable time windows enable multi-scale temporal analysis, balancing short-term events with long-term trends through flexible scaling.

### *3.3. Interactive Visualization for Trajectory Dynamic Topic*

To provide an intuitive dynamic analysis of trajectory topics, we develop an interactive visualization system. Figure 2 illustrates the user interface of the proposed system, which presents the dynamics of trajectory topics in both spatial and temporal dimensions. Specifically, the system consists of three main components: a trajectory topic map, a topic frequency trend view, and a topic word evolution view. We describe each component in detail below.

#### 3.3.1. Trajectory Topic Map

The interactive trajectory map view is designed to support the geospatial exploration and analysis of the semantic information embedded in trajectory topics. As illustrated in Figure 2A, this component visualizes the spatial distribution of trajectory topics. Specifically, after a user selects topics from a target time interval using the control panel on the left of the map view, the map view renders the corresponding GPS sequences onto the basemap. The GPS sequences represent the most frequent trajectories for the selected time interval and topic. Each sequence is rendered as a colored line, and the line color identifies the trajectory topic. To avoid visual clutter, our design follows Tufte's principle of maximizing the data–ink ratio. Consequently, we use a grayscale basemap to ensure that the colored trajectories stand out. For visual representation, each trajectory topic is assigned a distinct color, creating a clear visual hierarchy. Furthermore, the topic frequency, a quantitative value, is mapped to a sequential color scale within the same hue, where higher topic frequency corresponds to a darker and more saturated color. The map supports standard interactions such as panning, zooming, and filtering to facilitate interactive exploration.

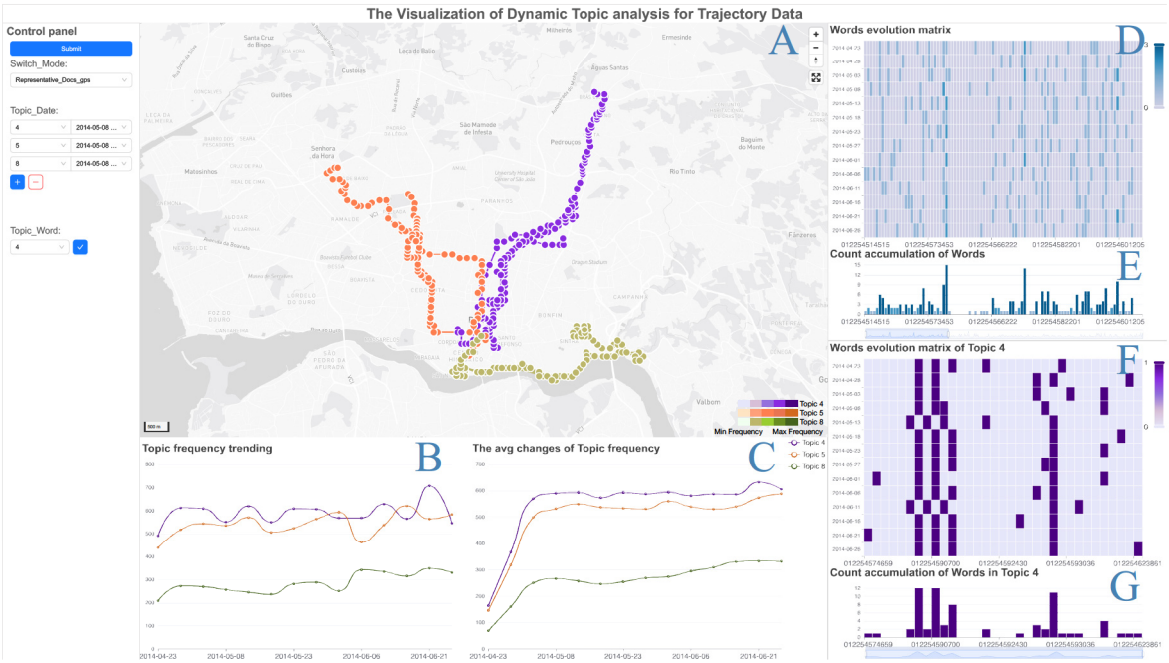



**Figure 2.** The overview of the proposed interactive visualization system. This system presents the visualization of topic analysis for trajectory data with three visual parts: trajectory topic map (component (**A**)), topic frequency trend (components (**B**,**C**)), and word evolution matrix (components (**D**–**G**)).

### 3.3.2. Topic Frequency Trend

To complement the spatial analysis of trajectory topics, we designed a topic frequency trend chart for revealing the temporal dynamics of different trajectory topics. This chart serves as a supplement to the map view, providing users with a linked component for coordinated spatial temporal analysis. To support the goals of identifying periodic patterns and comparing temporal dynamics, the chart's visual encoding is organized through the time dimension. Specifically, we compute the frequency of each topic over time to generate multiple time-series plots for topic frequency. Figure 2B displays the topic frequency trend as a line chart, where the x-axis represents time and the y-axis represents the topic frequency. Furthermore, to better illustrate the trend of frequency changes over time, we compute and plot the average changes in topic frequency, as shown in Figure 2C. To ensure consistency across views, the color encoding is kept consistent with that of the map view. Furthermore, to address the visual challenges of displaying multiple trend lines, we provide interactive features such as highlighting on mouse hover and displaying details on demand, which helps users focus on information of interest.

### 3.3.3. Topic Word Evolution

To refine the exploration granularity from the topic level to the individual topic word level, a topic word evolution view is proposed. The view supports analysis of the evolutionary trends of topic words over time. Specifically, the component is a composite view consisting of a temporal heatmap and an aggregated bar chart. These two coordinated views reveal both the overall dynamics and the detailed composition of topic word frequencies. Figure 2D,E present the evolution for all topic words, while Figure 2F,G illustrate the evolution for a single topic word. As shown in Figure 2D,F, the temporal frequency heatmap displays the frequency of each word within every time interval at fine granularity. Frequency values are encoded using a color channel. The heatmap is intended for identifying periodic and trend-based patterns and discerning co-occurrence relationships among

trajectory words. Figure 2E,G show the aggregated bar chart. The x-axis of the aggregated bar chart in Figure 2E is aligned with the x-axis of the heatmap in Figure 2D, with both axes representing topic words. The length of each bar in the aggregated bar chart encodes the total frequency of a topic word summed across all time intervals. The aggregated bar chart provides an overview of the frequency intensity for each topic word. The temporal heatmap and the aggregated bar chart are tightly coupled through strict alignment and a shared x-axis. Such coupling enables efficient visual correlation analysis. Additionally, the two views implement a classic focus and context design. The temporal frequency heatmap serves as the high-resolution focus view, displaying details of interest, while the underlying aggregated bar chart acts as a summarized context view. A focus and context design allows for a balanced exploration between microscopic details and macroscopic trends.

## 4. Experiments

### *4.1. Dataset and Setup*

The experiments were conducted on a server with an NVIDIA RTX 4090 GPU. The software environment included Python 3.10 and PyTorch 2.6. The dataset in this experiment includes Porto taxi trajectories [37]. This dataset contains 1,704,770 data records and records the taxi trajectory with GPS coordinates over one year from 1 July 2013 to 30 June 2014. Each record of the dataset comprises trip ID, a list of GPS coordinates, timestamp, and other types of attribute information. We employed the Porto dataset for the RoBERTa model with different training strategies, with training from scratch and fine-tuning. Basically, when training RoBERTa from scratch, we allocated 80% of the dataset records to training or fine-tuning the RoBERTa model, dividing this portion into 70% training, 15% validation, and 15% test sets. The remaining 20% served for trajectory dynamic topic modeling. For RoBERTa fine-tuning, we kept the same data partitioning strategy except for the training partition. The training partition was adapted to fine-tuning the RoBERTa model, while the validation and test partitions remained unchanged. We kept the same hyperparameters for the two training strategies.

### *4.2. Evaluation Metrics*

To evaluate the quality of trajectory topics, we employed multiple metrics, including coherence, diversity, stability, and distinctness. These metrics quantify the performance of dynamic topic modeling under various combination of components.

Topic coherence quantifies the interpretability of a trajectory topic [38]. For comparing the coherence of topics, we employed pointwise mutual information (PMI) for each term in a topic [39], which is defined as follows:

$$\mathrm{PMI}(w_i, w_j) = \log \frac{P(w_i, w_j)}{P(w_i) \cdot P(w_j)}. \tag{7}$$

The formulation presents the PMI of two words $w_i$ and $w_j$ from the same document. PMI quantifies term association within topics, providing an intuitive metric to evaluate BERTopic models. Topic coherence is calculated as follows:

$$\text{Coherence} = \frac{1}{N}\left(\frac{1}{N-1}\sum_{j \neq i} \mathrm{PMI}(w_i, w_j)\right). \tag{8}$$

Topic diversity is the proportion of distinct words among the first $N$ words of a topic. We evaluated the diversity of the top 10 words of the topic, and the topic diversity is defined as follows:

$$\text{Diversity} = \frac{\left|\bigcup_{t=0}^{K} W_t\right|}{(K+1) \times N}, \tag{9}$$

where $W_t = \{w_{t,1}, w_{t,2}, \ldots, w_{t,N}\}$ is the keywords for topic $t \in \{0, \cdots, K\}$.

Topic stability is adapted to measuring the degree of change in topic strength within a time window. Specifically, topic stability is represented as the average of the variance of topic frequencies across all time windows, and finally, the normalized stability score is calculated. Let $\{f_1, f_2, \ldots, f_T\}$ denote the frequency sequence of topic $k$ and $p$ denote the window size. The topic stability of topic $k$ in the $i$-th window is defined as $\sigma_i^2 = \frac{1}{w}\sum_{j=i}^{i+w-1}(f_j - \mu_i)^2$, where $\mu_i = \frac{1}{w}\sum_{j=i}^{i+w-1} f_j$ is the mean value in window $i$. The average variance of all windows is denoted by $\bar{\sigma}^2 = \frac{1}{T-w+1}\sum_{i=1}^{T-w+1}\sigma_i^2$. The normalized stability score of topic $k$ is calculated as $S_k = 1 - \frac{\bar{\sigma}^2}{\max(\{f_t\}_{t=1}^{T})^2}$, where $\bar{\sigma}$ represents the average standard deviation across all windows. The global topic stability $S_{\text{global}}$ can be calculated with the following formula:

$$\text{Stability} = \frac{1}{|K|}\sum_{k \in K} S_k. \tag{10}$$

### *4.3. Results and Analysis*

In this part, we evaluate the performance of trajectory topic modeling on the Porto dataset. Specifically, we first compare the topic modeling performance of the proposed method with traditional topic modeling methods. Then, we conduct ablation studies to determine how different parts of the model contribute to the proposed model. We quantify the influence of different components on the performance of the proposed method. Finally, we analyze the effects of different hyperparameters settings on the performance of the proposed method.

#### 4.3.1. Ablation Studies with Different Components

To validate the performance contribution of components including the trajectory embedding model and the representation model from adjustable dynamic topic modeling, we compare the performance of dynamic topic modeling with two training strategies, two tokenizers, and two representation models. The performance metrics include topic coherence, topic diversity, topic stability, and the number of topics. The performance results are shown in Table 1.

Topic coherence achieves the peak point at 3.546 for the component configuration of retraining strategy, WordLevel tokenizer, and KeyBERTInspired representation model. This demonstrates that the model captures adjacent movement patterns using domain-specific retraining and word-level tokenization. Notably, retraining from scratch yields significantly higher topic coherence metrics than fine-tuning the pretrained model. Specifically, with identical tokenizer and representation model settings, the retraining strategy achieves 40.16% higher average coherence scores than fine-tuning pretrained models. These results are competitive compared with traditional static models like LDA reported in previous studies [24].

**Table 1.** Execution time comparison of modeling dynamic topics of trajectories with different training strategies, tokenizers, and representation models.

| Training Strategy | Tokenizer | Representation Model | Coherence | Diversity | Stability | Topic Number |
|---|---|---|---|---|---|---|
| Fine-tuning | BPE | KeyBERTInspired | 2.497 ± 0.229 | 0.990 ± 0.022 | 0.987 ± 0.003 | 3.000 ± 0.000 |
| Fine-tuning | BPE | MaximalMarginalRelevance | 2.226 ± 0.118 | 1.000 ± 0.000 | 0.983 ± 0.001 | 3.000 ± 0.000 |
| Fine-tuning | WordLevel | KeyBERTInspired | 1.935 ± 0.477 | 0.982 ± 0.029 | 0.982 ± 0.001 | 3.600 ± 0.894 |
| Fine-tuning | WordLevel | MaximalMarginalRelevance | 2.449 ± 0.442 | 1.000 ± 0.000 | 0.982 ± 0.001 | 3.600 ± 0.894 |
| Retraining | BPE | KeyBERTInspired | 2.873 ± 0.296 | 1.000 ± 0.000 | 0.982 ± 0.000 | 3.000 ± 0.000 |
| Retraining | BPE | MaximalMarginalRelevance | 3.234 ± 0.002 | 1.000 ± 0.000 | 0.982 ± 0.001 | 3.000 ± 0.000 |
| Retraining | WordLevel | KeyBERTInspired | 3.546 ± 0.036 | 0.926 ± 0.015 | 0.988 ± 0.001 | 25.200 ± 1.100 |
| Retraining | WordLevel | MaximalMarginalRelevance | 2.862 ± 0.038 | 0.908 ± 0.009 | 0.988 ± 0.000 | 26.200 ± 0.837 |

The improvement suggest that retraining captures adjacent movement patterns of trajectories more effectively than pretraining.

Topic diversity exceeds 0.9 across all training component configurations. Our method uses a RoBERTa encoder and BERTopic for trajectory topic modeling. The high topic diversity scores demonstrate its effective topic separation. Our method uses a retraining strategy and a word-level tokenizer for dynamic topic modeling. However, it shows slightly lower topic diversity than other method combinations. This trade-off is acceptable. Our method produces finer-grained topics with the KeyBERTInspired representation model and achieves the best coherence scores.

Topic stability consistently exceeds 0.982, as shown in Table 1. This metric validates the temporal stability of our trajectory topics across time slices. Additionally, the model achieves the highest stability with the retraining strategy, WordLevel tokenizer, and KeyBERTInspired components. Moreover, the retraining strategy slightly improves topic stability, and the type of tokenizer has minimal impact.

The number of topics shows significant variation. As shown in Table 1, the model generates 25.2 topics for trajectory modeling with the retraining, WordLevel tokenizer, and KeyBERTInspired components. This is 740% higher than the fine-tuning training strategy configuration with three topics. The Porto dataset contains complex movement semantics. A small number of trajectory topics cannot fully cover all movement patterns. The difference in topic model quantity directly reflects our model's effectiveness in capturing trajectory topics. Additionally, with only the retraining strategy but keeping BPE tokenization, the number of topics is small. This shows that retraining alone is insufficient and requires combining with the WordLevel tokenizer.

#### 4.3.2. Hyperparameter Optimization for Components

To quantify the impact of different hyperparameters on the BERTopic model, this section compares diversity and stability results under various hyperparameter settings.

The first experiment evaluates the effect of the number of topic words on topic diversity. Topic diversity measures the uniqueness across different topics and is calculated as the proportion of unique words among the top N words of all topics. The experiment computes topic diversity scores for a topic word count ranging from 3 to 10.

Figure 3 illustrates the topic diversity scores for different numbers of topic words, revealing a slight downward trend as the word count increases. The diversity score reaches a peak of approximately 0.949 when using the top three words. As the number of included words increases to 10, the diversity score decreases to 0.923. The relationship is not strictly linear. The diversity score exhibits minor fluctuations. For example, when the word

count increases from five to six, the score rises from 0.923 to 0.936. However, the overall trend remains downward, suggesting that adding more topic words slightly reduces the collective diversity of the topic set. The highest-probability words in a topic are generally highly unique and representative, yielding a higher diversity value. In contrast, lower ranked words are often common or background terms prevalent across multiple topics. The inclusion of such common words reduces the proportion of unique terms among topics, thus causing a decrease in the diversity score.

Topic stability measures the degree of topic change over a specified time window. This experiment evaluates the differences in topic stability across various time-window lengths. Figure 4 presents a box plot of the topic stability score distribution for time windows ranging from 2 to 8. Figure 4 indicates a significant impact of window length on topic stability. The general trend shows a decrease in the median topic stability as the time-window size increases. For example, a window length of 2 yields a median stability score of 0.9937, the highest value among all tested settings. When the window length increases to 8, the median drops to 0.9894. This downward trend is particularly pronounced when the window length increases from 2 to 4, after which the decline becomes more gradual.

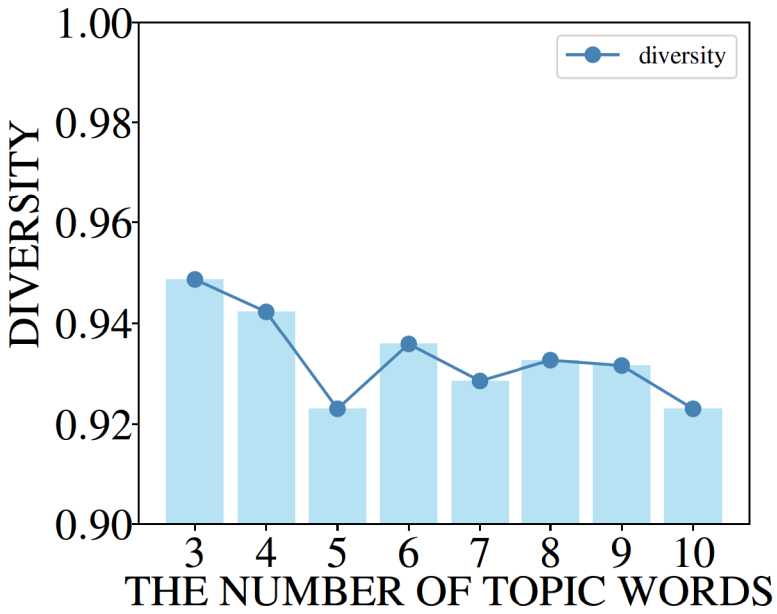



**Figure 3.** Topic diversity trend with increasing number of topic words.

In addition to the central tendency, data dispersion also increases with window length. For a window length of 2, the distribution of stability scores is the most concentrated, with an Interquartile Range (IQR) of 0.0073. This low IQR value indicates the highest result consistency. As the window length increases, data dispersion also increases. For instance, at a window length of 4, the IQR grows to 0.0111, and the minimum observed value drops from 0.9751 to 0.9544, for window lengths of 2 and 4, respectively. A larger window size, therefore, leads to greater result fluctuation and lower stability. In summary, smaller time windows compare adjacent time slices. The content and topic distributions of adjacent slices have naturally higher similarity, thus yielding higher stability scores. When the window expands, the analysis incorporates texts from a longer time span with potentially greater content variation. A longer time span leads to lower similarity between computed topic distributions and a lower stability score.

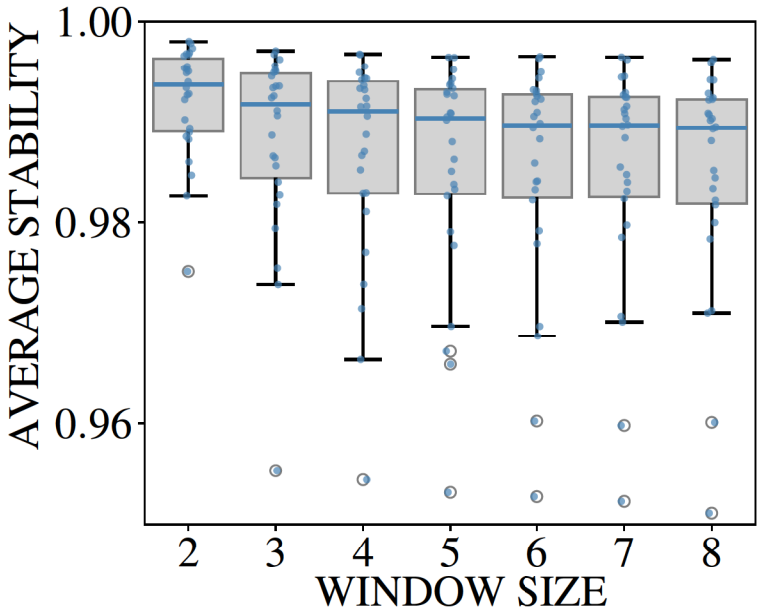



**Figure 4.** The box plot of topic stability with increasing window size.

### *4.4. Case Study*

#### 4.4.1. The Comparison of Trajectory Topics Across Time Slices

This case study demonstrates the effectiveness of the proposed visualization system for analyzing trajectory topic evolution by examining feature changes across adjacent time slices. Specifically, the analysis uses the proposed dynamic trajectory topic algorithm to compute dynamic topics from the Porto dataset. Figure 5 presents a comparison of the spatial distributions for the most frequent representative trajectories within trajectory topic 4, using a 5-day time interval. Trajectory topic 4 exhibits a spatial distribution from the southwest to the northeast, passing through commercial and residential areas, and primarily represents urban commute flows. As shown in Figure 5, the spatial distribution of topic 4 appears generally stable across different time slices.

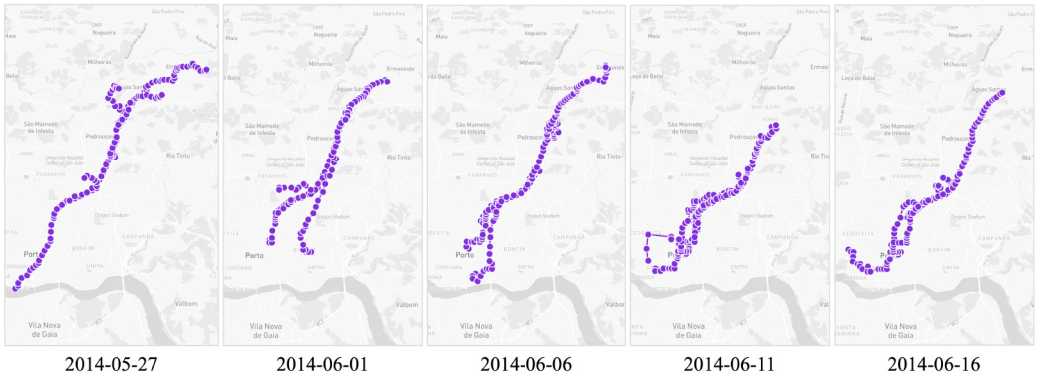



**Figure 5.** Spatial distribution of trajectory topic 4 across adjacent time slices.

The key dynamic change within topic 4 manifests as a difference in commute distance between weekdays and non-workdays. Specifically, on weekdays (e.g., 27 May 2014 and 6 June 2014), the trajectories of topic 4 extend significantly farther to the southwest than trajectories on non-workdays (e.g., 1 June 2014). In summary, the visualized changes in the spatial distribution of representative trajectories for topic 4 clearly reveal the topic's dynamic evolution. This clear presentation reflects the effectiveness of the proposed method in modeling dynamic trajectory topics.

#### 4.4.2. The Comparison of Different Trajectory Topics

This case study compares the details of different trajectory topics by examining differences in spatial distribution. The analysis again uses the proposed dynamic topic modeling method on the Porto dataset to compare trajectory topic 4 and trajectory topic 5.

First, a spatial comparison reveals distinct patterns. Figure 6 shows the spatial distributions for the most frequent representative trajectories of topic 4 and topic 5 within the same time slice. The orange trajectories, representing topic 5, exhibit regional clustering near tourist attractions (e.g., museums and squares), restaurants, and accommodations. The spatial flow of topic 5 radiates from a central accommodation area, primarily representing tourism flows. As shown in Figure 6, the overall difference in spatial distribution between topic 4 and topic 5 is clearly visible. For a more direct comparison, Figure 7 plots the representative topic words for both topics. Representative topic words denote high-frequency GPS points. The map in Figure 7 illustrates that the topic words for topic 4 are distributed along a southwest-to-northeast commute corridor, whereas the topic words for topic 5 are clustered around the tourist areas.

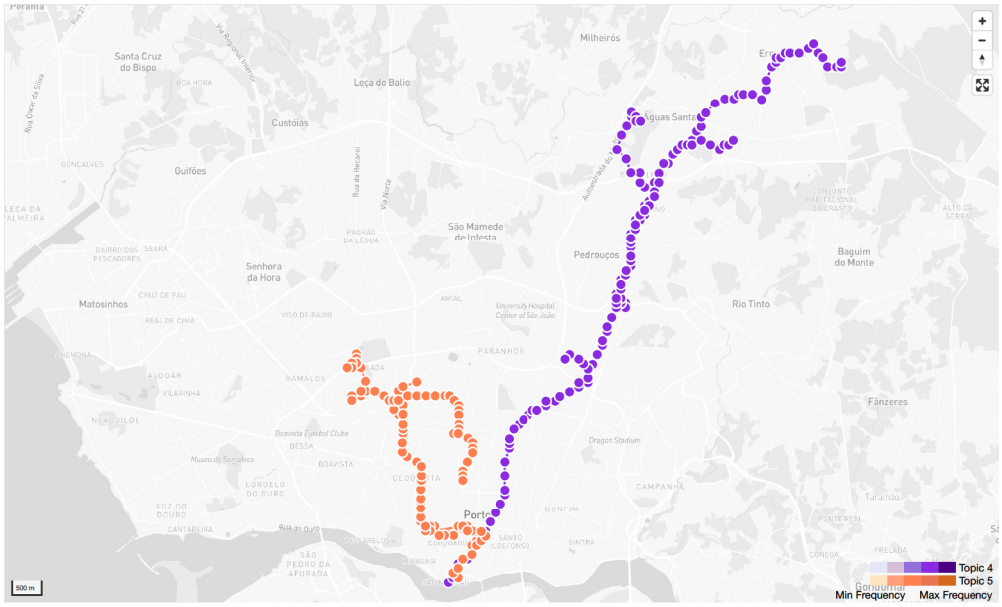



**Figure 6.** The comparison of spatial distribution of trajectories. The purple line and orange line in the map present the trajectories of topic 4 and topic 5, respectively.

Second, a statistical frequency analysis provides further insights. Figure 8 presents the frequency trends for topic 4 and topic 5. The chart in Figure 8a shows that the frequency of topic 4 is higher than the frequency of topic 5 in 85.71% of the time intervals. This observation suggests that the commute flow from topic 4 constitutes a larger portion of the city's overall trajectory flow than the tourism flow from topic 5. Additionally, Figure 8b displays the rate of frequency change. In this chart, the frequency change line for topic 4 is consistently above the line for topic 5, indicating a higher rate of change for topic 4 within each time slice.

In conclusion, although topic 4 and topic 5 show similar frequency trends at a statistical level, the two topics possess markedly different spatial distributions. Therefore, significant differences exist between distinct trajectory topics even within the same time slice. This finding reflects the effectiveness of the proposed method in establishing and analyzing trajectory topics.

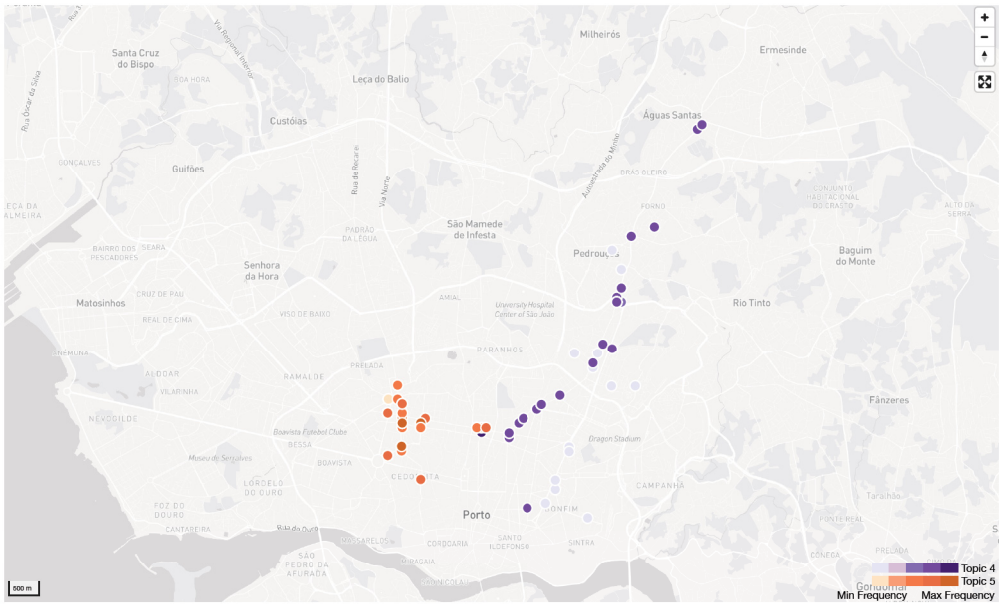
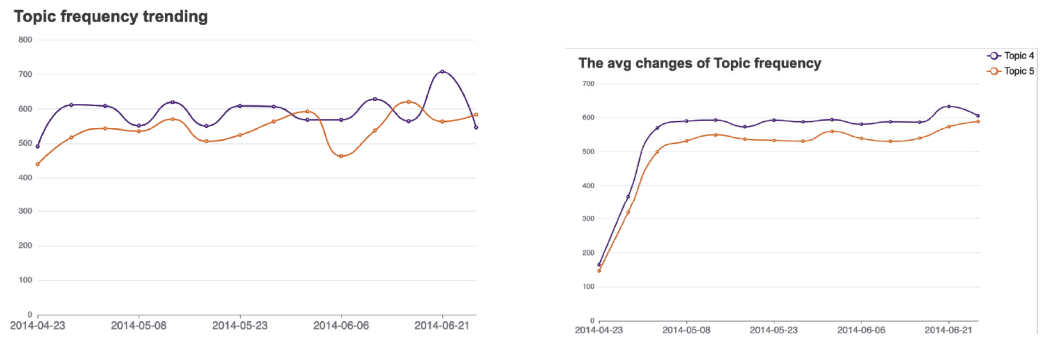



**Figure 7.** The spatial distribution of trajectory topic 4 and trajectory topic 5. The purple points and orange points in the map present the high-frequency topic words of topic 4 and topic 5, respectively.




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

**Figure 8.** The frequency trend of trajectory topics. The purple line represents topic 4, and the orange line represents topic 5. (**a**) The topic frequency trends of topic 4 and topic 5. (**b**) The change in frequencies for topic 4 and topic 5 over several time intervals.

## 5. Discussion

The proposed method holds significant value for the analysis of trajectory data. By combining a domain-specific language model with a flexible topic modeling method, this work provides a powerful analytical tool for discovering dynamic movement patterns from raw GPS sequences. The method has substantial potential impact across multiple domains. Urban planners and traffic engineers can use the system to understand the evolution of commute patterns and identify traffic characteristics. In tourism and commerce, the approach enables the identification of changing visitor flows and the analysis of activity hotspots around commercial centers. Furthermore, the integration of such analytics with IoT platforms could further transform automation in these sectors [40].

Furthermore, the interactive visualization system allows stakeholders without extensive data science expertise to perform complex spatio-temporal analysis, bridging the gap between advanced data mining and practical decision making. The unsupervised nature of the method also reduces dependency on manually labeled data, increasing the applicability of the method to diverse trajectory datasets.

While the proposed method shows strong performance, several areas offer opportunities for improvement. First, the current trajectory representation relies solely on geospatial

coordinates encoded via Morton codes. Future work could incorporate richer contextual information and multimodal data, such as timestamps or external factors, to obtain more specific topic representations and to add greater semantic depth to the generated topics. Second, the evaluation is conducted on a single large-scale taxi trajectory dataset from Porto. Validating the method on diverse trajectory datasets is necessary to ensure generalization. The characteristics of taxi movement, being demand-driven, may differ from other forms of mobility like public transportation, pedestrian movement, or logistics. Extending and validating the method on datasets from these other domains remains an important direction for future research. Finally, training the RoBERTa model is computationally intensive. Future work will investigate strategies like model pruning or quantization to reduce training costs. Additionally, we plan to use optimization algorithms, such as MFO, SOA, or HBA, to tune hyperparameters for better efficiency. Future work should, therefore, consider the trade-off between performance and efficiency for practical implementations.

## 6. Conclusions

This paper presents a complete method for the modeling, analysis, and visualization of dynamic topics in trajectory data. A core contribution of this work is the development of a dynamic topic model. The model uses domain-specific RoBERTa embeddings and a BERTopic-based pipeline to overcome the limitations of static methods and to capture the complex spatial context of movement patterns. The proposed method effectively identifies semantically coherent topics from raw GPS sequences. Another key contribution is the incorporation of an adjustable time window. This feature provides flexibility for analyzing trajectory dynamics across multiple temporal scales. Finally, the work introduces an interactive visualization system with spatio-temporal views. The system successfully transforms abstract model outputs into an intuitive format, supporting the direct exploration and clear understanding of how trajectory topics evolve across space and time. The entire method provides an effective solution for dynamic trajectory topic analysis. Future research will focus on two main directions: (1) integrating multimodal data, such as timestamps and POIs, to enrich trajectory semantics and (2) employing model compression strategies and optimization algorithms to improve computational efficiency for large-scale applications.

**Author Contributions:** Conceptualization, H.C. and Y.W.; software, J.L. (Jing Lei) and G.W.; data preparation, G.W. and J.L. (Jing Liao); evaluation, H.C., W.Z., and F.W.; manuscript, H.C. and G.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research study was funded by Key Research and Development Program of Sichuan Province under grant number 18ZS2152, Doctoral Research Fund of Southwest University of Science and Technology under grant number 25ZX7125, and University-Industry Collaborative Education Program under grant number 219-25SJJG21.

**Data Availability Statement:** Data are available online [37].

**Conflicts of Interest:** Author Jing Lei was employed by the Technical Center, Mianyang Xinchen Engine 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. Hu, D.; Chen, L.; Fang, H.; Fang, Z.; Li, T.; Gao, Y. Spatio-temporal trajectory similarity measures: A comprehensive survey and quantitative study. *IEEE Trans. Knowl. Data Eng.* **2023**, *36*, 2191–2212. [CrossRef]
2. Wang, S.; Bao, Z.; Culpepper, J.S.; Cong, G. A survey on trajectory data management, analytics, and learning. *Acm Comput. Surv. (CSUR)* **2021**, *54*, 1–36. [CrossRef]
3. Sherstinsky, A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Phys. D Nonlinear Phenom.* **2020**, *404*, 132306. [CrossRef]

4. Fu, T.Y.; Lee, W.C. Trembr: Exploring road networks for trajectory representation learning. *ACM Trans. Intell. Syst. Technol. (TIST)* **2020**, *11*, 1–25. [CrossRef]
5. Yang, S.B.; Guo, C.; Hu, J.; Tang, J.; Yang, B. Unsupervised path representation learning with curriculum negative sampling. *arXiv* **2021**, arXiv:2106.09373. [CrossRef]
6. Yang, S.B.; Guo, C.; Hu, J.; Yang, B.; Tang, J.; Jensen, C.S. Weakly-supervised temporal path representation learning with contrastive curriculum learning. In Proceedings of the 2022 IEEE 38th International Conference on Data Engineering (ICDE), Kuala Lumpur, Malaysia, 9–12 May 2022; IEEE: New York, NY, USA, 2022; pp. 2873–2885.
7. Liu, H.; Han, J.; Fu, Y.; Li, Y.; Chen, K.; Xiong, H. Unified route representation learning for multi-modal transportation recommendation with spatiotemporal pre-training. *VLDB J.* **2023**, *32*, 325–342. [CrossRef]
8. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30* , 6000–6010.
9. Chen, Y.; Li, X.; Cong, G.; Bao, Z.; Long, C.; Liu, Y.; Chandran, A.K.; Ellison, R. Robust road network representation learning: When traffic patterns meet traveling semantics. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management, Queensland, Australia, 1–5 November 2021; pp. 211–220.
10. Mao, Z.; Li, Z.; Li, D.; Bai, L.; Zhao, R. Jointly contrastive representation learning on road network and trajectory. In Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, 17–21 October 2022; pp. 1501–1510.
11. Yang, S.B.; Hu, J.; Guo, C.; Yang, B.; Jensen, C.S. Lightpath: Lightweight and scalable path representation learning. In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Long Beach, CA, USA, 6–10 August 2023; pp. 2999–3010.
12. Wu, C.; Xiang, L.; Chen, L.; Zhong, Q.; Wu, X. Learning universal trajectory representation via a siamese geography-aware transformer. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 64. [CrossRef]
13. Zhou, S.; Shang, S.; Chen, L.; Jensen, C.S.; Kalnis, P. RED: Effective Trajectory Representation Learning with Comprehensive Information. *Proc. VLDB Endow.* **2024**, *18*, 80–92. [CrossRef]
14. Blei, D.M.; Lafferty, J.D. Dynamic topic models. In Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA, USA, 25 June 2006; pp. 113–120.
15. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
16. Deng, Z.; Weng, D.; Liu, S.; Tian, Y.; Xu, M.; Wu, Y. A survey of urban visual analytics: Advances and future directions. *Comput. Vis. Media* **2023**, *9*, 3–39. [CrossRef] [PubMed]
17. Zhao, W.; Wang, G.; Wang, Z.; Liu, L.; Wei, X.; Wu, Y. A uncertainty visual analytics approach for bus travel time. *Vis. Inform.* **2022**, *6*, 1–11. [CrossRef]
18. Zhao, W.; Jiang, H.; Tang, K.; Pei, W.; Wu, Y.; Qayoom, A. Knotted-line: A visual explorer for uncertainty in transportation system. *J. Comput. Lang.* **2019**, *53*, 1–8. [CrossRef]
19. Chu, D.; Sheets, D.A.; Zhao, Y.; Wu, Y.; Yang, J.; Zheng, M.; Chen, G. Visualizing hidden themes of taxi movement with semantic transformation. In Proceedings of the 2014 IEEE Pacific Visualization Symposium, Yokohama, Japan, 4–7 March 2014; IEEE: New York, NY, USA, 2014; pp. 137–144.
20. Grootendorst, M.R. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv* **2022**, arXiv:2203.05794. Available online: https://api.semanticscholar.org/CorpusID:247411231 (accessed on 11 February 2025).
21. Lee, D.; Seung, H.S. Algorithms for non-negative matrix factorization. *Adv. Neural Inf. Process. Syst.* **2000**, *13*, 535–541.
22. Wallach, H.M. Topic modeling: Beyond bag-of-words. In Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA, USA, 25 June 2006; pp. 977–984.
23. Liao, L.; Wu, J.; Zou, F.; Pan, J.; Li, T. Trajectory topic modelling to characterize driving behaviors with GPS-based trajectory data. *J. Internet Technol.* **2018**, *19*, 815–824.
24. Huang, L.; Wen, Y.; Guo, W.; Zhu, X.; Zhou, C.; Zhang, F.; Zhu, M. Mobility pattern analysis of ship trajectories based on semantic transformation and topic model. *Ocean. Eng.* **2020**, *201*, 107092. [CrossRef]
25. Liu, D.; Xu, P.; Ren, L. TPFlow: Progressive partition and multidimensional pattern extraction for large-scale spatio-temporal data analysis. *IEEE Trans. Vis. Comput. Graph.* **2018**, *25*, 1–11. [CrossRef]
26. Liu, H.; Jin, S.; Yan, Y.; Tao, Y.; Lin, H. Visual analytics of taxi trajectory data via topical sub-trajectories. *Vis. Inform.* **2019**, *3*, 140–149. [CrossRef]
27. Tao, Y.; Tang, Y. Progressive visual analysis of traffic data based on hierarchical topic refinement and detail analysis. *J. Vis.* **2023**, *26*, 367–384. [CrossRef]
28. Liu, L.; Zhan, H.; Liu, J.; Man, J. Visual analysis of traffic data via spatio-temporal graphs and interactive topic modeling. *J. Vis.* **2019**, *22*, 141–160. [CrossRef]
29. Mohammadiha, N.; Smaragdis, P.; Panah, G.; Doclo, S. A State-Space Approach to Dynamic Nonnegative Matrix Factorization. *IEEE Trans. Signal Process.* **2015**, *63*, 949–959. [CrossRef]

30. Yao, D.; Zhang, C.; Zhu, Z.; Huang, J.; Bi, J. Trajectory clustering via deep representation learning. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; IEEE: New York, NY, USA, 2017; pp. 3880–3887.
31. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv* **2019**, arXiv:1907.11692. Available online: http://arxiv.org/abs/1907.11692 (accessed on 11 February 2025).
32. Mielke, S.J.; Alyafeai, Z.; Salesky, E.; Raffel, C.; Dey, M.; Gallé, M.; Raja, A.; Si, C.; Lee, W.Y.; Sagot, B.; et al. Others. Between words and characters: A brief history of open-vocabulary modeling and tokenization in nlp. *arXiv* **2021**, arXiv:2112.10508.
33. He, J.; Chen, C. Spatio-temporal analytics of topic trajectory. In Proceedings of the 9th International Symposium on Visual Information Communication and Interaction, Dallas, TX, USA, 24–26 September 2016; pp. 112–116.
34. Gao, X.; Liao, C.; Chen, C.; Li, R. Visual exploration of cycling semantics with GPS trajectory data. *Appl. Sci.* **2023**, *13*, 2748. [CrossRef]
35. Morton, G.M. *A Computer Oriented Geodetic Data Base and a New Technique in File Sequencing*; International Business Machines Company: Armonk, NY, USA, 1966.
36. Rahman, M.F.; Liu, W.; Suhaim, S.B.; Thirumuruganathan, S.; Zhang, N.; Das, G. Hdbscan: Density based clustering over location based services. *arXiv* **2016**, arXiv:1602.03730. [CrossRef]
37. Cross, C. Taxi Trajectory Data. Available online: https://www.kaggle.com/datasets/crailtap/taxi-trajectory/data (accessed on 11 February 2025).
38. Mimno, D.; Wallach, H.; Talley, E.; Leenders, M.; McCallum, A. Optimizing semantic coherence in topic models. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, Edinburgh, UK, 27–31 July 2011; pp. 262–272.
39. Aletras, N.; Stevenson, M. Evaluating topic coherence using distributional semantics. In Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)–Long Papers, Potsdam, Germany, 19–22 March 2013; pp. 13–22.
40. Addula, S.R.; Tyagi, A.K.; Naithani, K.; Kumari, S. Blockchain-empowered Internet of things (IoTs) platforms for automation in various sectors. *Artif.-Intell.-Enabled Digit. Twin Smart Manuf.* **2024**, 201–217. [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.

*electronics*

MDPI

*Article*

# Enhanced Semantic BERT for Named Entity Recognition in Education

**Ping Huang *, Huijuan Zhu, Ying Wang, Lili Dai and Lei Zheng**

College of Computer and Artificial Intelligence, Nanjing University of Science and Technology Zijin College, Nanjing 210023, China; zhuhuijuan@njust.edu.cn (H.Z.); wangying826@njust.edu.cn (Y.W.); dailili992@njust.edu.cn (L.D.); zhenglei@hhu.edu.cn (L.Z.)

* Correspondence: huangping984@njust.edu.cn; Tel.: +86-025-85786129

**Abstract**

To address the technical challenges in the educational domain named entity recognition (NER), such as ambiguous entity boundaries and difficulties with nested entity identification, this study proposes an enhanced semantic BERT model (ES-BERT). The model innovatively adopts an education domain, vocabulary-assisted semantic enhancement strategy that (1) applies the term frequency–inverse document frequency (TF-IDF) algorithm to weight domain-specific terms, and (2) fuses the weighted lexical information with character-level features, enabling BERT to generate enriched, domain-aware, character–word hybrid representations. A complete bidirectional long short-term memory-conditional random field (BiLSTM-CRF) recognition framework was established, and a novel focal loss-based joint training method was introduced to optimize the process. The experimental design employed a three-phase validation protocol, as follows: (1) In a comparative evaluation using 5-fold cross-validation on our proprietary computer-education dataset, the proposed ES-BERT model yielded a precision of 90.38%, which is higher than that of the baseline models; (2) Ablation studies confirmed the contribution of domain-vocabulary enhancement to performance improvement; (3) Cross-domain experiments on the 2016 knowledge base question answering datasets and resume benchmark datasets demonstrated outstanding precision of 98.41% and 96.75%, respectively, verifying the model's transfer-learning capability. These comprehensive experimental results substantiate that ES-BERT not only effectively resolves domain-specific NER challenges in education but also exhibits remarkable cross-domain adaptability.

**Keywords:** NER; BERT; semantic enhancement; BiLSTM; CRF; focal loss

## 1. Introduction

Named entity recognition (NER) in the educational domain serves as a core natural language processing (NLP) application, providing crucial support for educational informatization by automatically identifying key entities—such as subject concepts, institutions, and teaching resources—in educational texts [1,2]. By extracting structured information from unstructured data, NER supports advanced applications such as relation extraction [3], question answering [4], and knowledge graph construction [5]. Chinese educational NER faces specific challenges due to linguistic characteristics. The inherent ambiguity of Chinese word boundaries necessitates simultaneous attention to word segmentation and entity detection. Furthermore, Chinese named entities often exhibit a coarser granularity than words, ranging from single characters to multi-character compounds. Consequently, ac-

curate boundary identification and enriched lexical semantics are critical for improving model performance [6,7].

The prevalence of specialized terminology, abbreviations, and compound words in the Chinese educational domain gives rise to significant challenges, including entity boundary ambiguity and nested entity structures. The phrase "University Physics," for example, represents an indivisible course entity yet is susceptible to misinterpretation as a generic descriptive phrase. Furthermore, compound terms such as "Artificial Intelligence Python Practice" exhibit nesting, where the core entity contains subordinate entities denoting a field and a pedagogical method, respectively. The accurate resolution of boundary ambiguity necessitates precise contextual comprehension, whereas the identification of nested entities requires capabilities for multi-granular semantic recognition. BERT provides a transformative solution through its bidirectional Transformer architecture, which acquires dynamic, context-sensitive word representations via deep pre-training, thereby facilitating nuanced boundary detection and hierarchical analysis. The model's multi-layer attention mechanism additionally enables the learning of features at varying granularities, offering inherent support for nested entity recognition. Together, these capabilities establish a consolidated and powerful semantic modeling foundation for addressing ambiguity and structural complexity in specialized domains.

To address the challenges of ambiguous boundaries and nested entity recognition in the educational domain NER, we propose an enhanced semantic BERT model (ES-BERT). By fusing domain-specific lexical knowledge from educational dictionaries with character-level features, ES-BERT enriches semantic representations. Experiments demonstrate that this fusion strategy mitigates semantic gaps and boundary ambiguity in Chinese character-level NER, achieving improvement in F1 score over baseline models. The main contributions of this study are threefold:

- Model Architecture: Building upon BERT, we propose ES-BERT (enhanced semantic BERT), a hybrid-fusion model that combines domain-specific lexical features with character-level representations. This dual-path design synergistically enhances granular semantic encoding, providing a theoretical foundation for educational NER performance improvement.
- Training Objective: To address class imbalance (non-entity token dominance) and boundary detection challenges, we integrate a novel focal loss-based joint training objective. This approach optimizes label constraints and boundary representations, significantly improving accuracy for boundary-sensitive entities.
- Resource Construction: We manually curate a computer education corpus with 1000 expert-annotated research abstracts, processed through tokenization and rigorous annotation protocols. Additionally, we compile a domain-specific dictionary to mitigate the current lack of linguistic resources in this field.

## 2. Related Work

The rapid advancement of named entity recognition (NER) has led the research community to favor hybrid architectures that synergistically combine neural networks with statistical learning. A prime example is the BERT+BiLSTM+CRF framework, which accomplishes efficient sequence labeling through a multi-stage complementary process: BERT generates contextually rich character-level embeddings; BiLSTM captures long-range bidirectional dependencies to enhance feature coherence; and CRF optimizes the label sequence globally by modeling transition constraints, thereby preventing invalid predictions. By integrating deep representation learning with structured prediction, this framework exhibits robust adaptability and stability across diverse NER domains. Dong et al. [8] proposed BERT-WWM-EXT, which incorporates Whole Word Masking during

pre-training to enhance hierarchical representation learning. When integrated with BiLSTM-CRF, this approach demonstrated superior boundary detection capabilities in judicial texts. Hu et al. [9] proposed a BERT-BiLSTM-CRF-based named entity recognition method for core educational technology courses to enhance learning efficiency, reporting improved performance over conventional CRF and BiLSTM-CRF models. Xie et al. [10] enhanced the BERT-BiLSTM-CRF paradigm with contextual-semantic enhancement layers, setting new benchmarks on the MSRA and People's Daily corpora, particularly for polysemous entities. During the ongoing evolution of the BERT model, a number of efficient variants have emerged—represented by RoBERTa [11] and ALBERT [12]—which have been widely applied in named entity recognition tasks.

With the advancement of research, semantic enhancement techniques have demonstrated effectiveness in named entity recognition (NER). Numerous studies have explored the integration of character-level, word-level, and semantic information to enhance the performance of Chinese NER tasks. Chen et al. [13] aggregated multiple word-level features associated with characters and incorporated them into BERT for text feature enhancement, successfully addressing Chinese nested named entity recognition. Liu et al. [14] proposed LEBERT, a lexically enhanced BERT model for Chinese sequence labeling. The model integrates external lexical knowledge directly into BERT layers through a lexical adapter and performs deep lexical knowledge fusion, achieving state-of-the-art performance in NER and related tasks. Wu et al. [15] developed an entity recognition model that incorporates external dictionaries for feature enhancement and adversarial training. Their approach constructs character–word pairs from dictionaries, employs a dictionary adapter module for feature fusion, and utilizes FGM (Fast Gradient Method) to enhance model robustness. Sheng et al. [16] designed a lexicon-enhanced BiLSTM-CRF model that combines character and word embeddings at the embedding layer to enrich initial text representations. These enhanced embeddings serve as input to a BiLSTM network, with the CRF layer performing final label decoding. Wang et al. [17] introduced an interactive fusion method for integrating character and word information in Chinese NER. Their method employs an interactive graph structure to merge character and lexical features, resulting in more comprehensive feature representations. To more clearly illustrate the technical approaches of existing semantic enhancement techniques, a summary is provided in Table 1.

**Table 1.** Research methods for semantic enhancement in Chinese NER.

| Method Category | Core Methodology | Reported Performance |
| --- | --- | --- |
| swM [13] | Integrates character-related lexical information and uses an MLP classifier to predict the start and end characters of spans for boundary detection. | Improved recognition accuracy in experiments on public datasets. |
| LEBERT-CRF [14] | Integrates external lexical knowledge into BERT layers through a designed lexicon adapter layer for deep fusion. | Achieved SOTA results on multiple Chinese NER benchmarks. |
| LEBERT-BiLSTM-CRF [15] | Fuses character and word vector features via a lexicon adapter module and uses FGM to enhance model robustness. | Improved model robustness and recognition performance. |
| BERT+meanword & BERT+softword [16] | Two lexicon-enhanced vector fusion methods: soft word frequency and average word vector. | Increased F1 score on two datasets. |
| BERI-SoftLexical-GAT-CRF [17] | Employs an interactive graph structure to integrate character and lexical information for more comprehensive feature representation. | F1 score surpassed previous best models on multiple datasets. |

Prior research has demonstrated that semantic enhancement techniques can effectively identify entity boundaries and extract rich semantic information from textual data, yielding measurable improvements in NER performance. To address the research gap in the educational domain NER, we propose ES-BERT, a novel architecture that integrates domain-adapted lexical features with character-level embeddings. Specifically, ES-BERT employs an educational domain lexicon to generate hybrid character–word representations augmented with

domain-specific semantics. The proposed framework simultaneously addresses two key challenges in educational NER, as follows: (1) ambiguous entity boundaries and (2) nested entity recognition difficulties, resulting in substantial performance gains.

## 3. ES-BERT with Enhanced Semantic Representation

The BERT model has demonstrated remarkable advantages in named entity recognition (NER) tasks, with its contextual understanding capabilities and fine-tuning flexibility establishing it as a state-of-the-art approach [18–20]. As a character-level model, BERT processes each Chinese character as an independent token, thus circumventing word segmentation errors. However, this architecture inherently lacks lexical information utilization—a critical component for capturing inter-character relationships and facilitating boundary detection [21].

To address this limitation, we propose ES-BERT (enhanced semantic BERT, Figure 1), which integrates word-level features with character representations to strengthen inter-character associations. This hierarchical semantic enhancement provides two key benefits, as follows: (1) more precise entity boundary identification and (2) richer contextual semantic extraction, which collectively advance the performance of Chinese NER. The construction process of the ES-BERT model is shown in Algorithm 1.

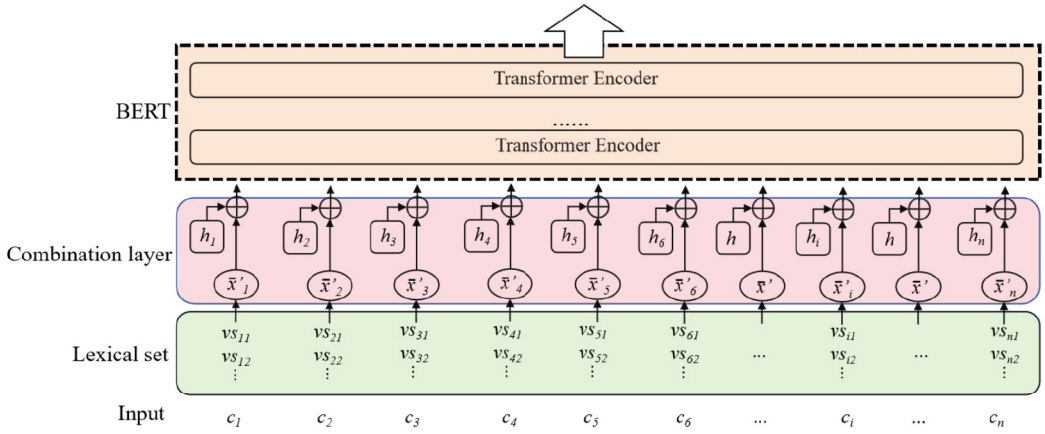



**Figure 1.** ES-BERT model.

### *3.1. Domain Lexicon Construction*

In the Chinese linguistic context, multi-character lexical units (versus single characters) inherently embody richer semantic priors and provide more definitive boundary indicators for named entity recognition. This phenomenon is particularly pronounced in specialized domains. Domain-specific terminologies exhibit strong predictive power that can effectively guide and refine model predictions.

To optimize lexical semantic representation, our approach incorporates two key components, as follows: (1) core computer science terminologies and (2) domain-specific vocabulary from both tertiary and vocational education contexts. The educational domain lexicon was compiled by selectively extracting relevant terms from Tencent's open-source Chinese word embeddings, focusing particularly on computer science terminology.

**Algorithm 1** Enhanced-semantic-BERT Chinese NER algorithm

**Input:** Chinese sentence $S = \{c_1, c_2, \ldots, c_n\}$, domain lexicon $D$
**Output:** Encoded semantic vector sequence $P = [p_1, p_2, \ldots, p_n]$

**Phase 1: Character–Lexicon Matching**
**for** $i = 1$ **to** $n$ **do**
  $vs_i \leftarrow \emptyset$ # *Initialize lexicon set for character* $c_i$
  Generate word combinations centered at $c_i$ in sentence context
  **for** each word combination $w$ **do**
    **if** $w \in D$ # *Verify match in domain lexicon* **then**
      $vs_i \leftarrow vs_i \cup \{w\}$
    **end if**
  **end for**
**end for**
$Scs \leftarrow \{(c_1, vs_1), (c_2, vs_2), \ldots, (c_n, vs_n)\}$
**Phase 2: Character–Word Vector Fusion**
**for** $i = 1$ **to** $n$ **do**
  $h_i \leftarrow \text{BERT_Embedding}(c_i)$ # *Character vector: token + segment + position embeddings*
  $x_i \leftarrow \emptyset$ # *Initialize word vector set*
  **for** each word $w_j \in vs_i$ **do**
    $x_{ij} \leftarrow \text{GetWordEmbedding}(e_j)$ # *Get word vector from domain lexicon*
    $x_i \leftarrow x_i \cup \{x_{ij}\}$
  **end for**
  **TF-IDF Weight Calculation**
  **for** each word vector $x_{ij} \in x_i$ **do**
    $w_{ij} \leftarrow TF(vs_{ij}) \times IDF(vs_{ij})$ # *Calculate TF-IDF weights*
  **end for**
  **Weighted Word Vector Fusion**
  $\bar{x}_i \leftarrow \mathbf{0}$ # *Initialize fused word vector*
  **for** $j = 1$ **to** $|x_i|$ **do**
    $\bar{x}_i \leftarrow \bar{x}_i + w_{ij} \times x_{ij}$
  **end for**
  **Dimension Alignment Transformation**
  $\bar{x}'_i \leftarrow W_2 \times \text{ReLU}(W_1 \times \bar{x}_i + b_1) + b_2$
  **Final Fusion**
  $h'_i \leftarrow h_i + \bar{x}'_i$ # *Obtain character–word fused vector*
**end for**
**Phase 3: BERT Encoding**
$H \leftarrow [h'_1, h'_2, \ldots, h'_n]$ # *Construct input sequence*
**Self-Attention Mechanism**
**for** each Transformer layer **do**
  $Q \leftarrow \text{Linear}(H) \times W^Q$
  $K \leftarrow \text{Linear}(H) \times W^K$
  $V \leftarrow \text{Linear}(H) \times W^V$
  $\text{Attention} \leftarrow \text{Softmax}(\frac{Q \times K^T}{\sqrt{d_k}}) \times V$
  **Multi-Head Attention**
  $\text{MultiHead} \leftarrow \text{Concat}(\text{head}_1, \ldots, \text{head}_h) \times W^O$
  **where** $\text{head}_i = \text{Attention}(Q_i, K_i, V_i)$
  $H \leftarrow \text{LayerNorm}(H + \text{MultiHead})$
  $H \leftarrow \text{LayerNorm}(H + \text{FFN}(H))$ # *Feed-forward network*
**end for**
$P \leftarrow H$ # *Output encoded semantic vector sequence*
**return** $P$

Leveraging this lexicon, we identify character-position-aware lexical candidates within input texts. This design intentionally circumvents error-prone word segmentation, since segmentation precision constitutes a fundamental bottleneck in Chinese NER systems.

Crucially, any segmentation errors would propagate through the recognition cascade, severely compromising entity identification reliability.

To address this fundamental challenge, our framework dynamically constructs domain-specific lexical candidate sets through comprehensive sentence-to-lexicon matching. The core algorithm operates on a character-by-character basis, as follows: For each Chinese character in the input sequence, it evaluates all possible contextual combinations. When a combination matches a domain lexicon entry, the corresponding word is incorporated into the character's lexical candidate set.

Formally, let $S = \{c_1, c_2, \ldots, c_n\}$ denote an n-character Chinese sentence. Our method generates a character–word sequence $S_{cs} = \{(c_1, vs_1), (c_2, vs_2), \ldots, (c_n, vs_n)\}$, where each $vs_i$ represents the potential lexical set containing all domain words that (1) include character $c_i$, (2) occur in sentence $S$, and (3) exist in the domain lexicon (see Figure 1). This is implemented using an optimized maximum matching algorithm with $O(n^2)$ time complexity.

### *3.2. Character–Word Vector Fusion*

Upon acquiring the character–word sets $(c_i, vs_i)$, we transform both characters and their associated words into vector representations $(h_i, x_i)$. Specifically, the character vector $h_i$ at position $i$ is derived from BERT's output representation, which aggregates three components, as follows: token embedding, segment embedding, and position embedding. The word vector set $x_i = \{x_{i1}, x_{i2}, \ldots, x_{im}\}$ contains all pre-trained lexical embeddings corresponding to character $c_i$, where each $x_{ij}$ is retrieved from our domain-specific lexicon.

To mitigate segmentation errors, we compute TF-IDF weights for candidate words, assigning lower weights to the following: (1) infrequent terms, and (2) out-of-vocabulary words. The weight $w_{ij}$ for the $j$-th word associated with character $c_i$ is calculated as follows:

$$w_{ij} = TF(vs_{ij}) * IDF(vs_{ij}) \tag{1}$$

where $TF(vs_{ij})$ represents the term frequency (TF) of word $vs_{ij}$ (the $j$-th word corresponding to character $c_i$) in the document, $IDF(vs_{ij})$ represents the inverse document frequency (IDF) of word $vs_{ij}$.

For each character $c_i$, we compute the weighted sum of all its corresponding word vectors to generate the fused lexical vector $\bar{x}_i$:

$$\bar{x}_i = \sum_{j=1}^{m} w_{ij} x_{ij} \tag{2}$$

To address the dimensionality mismatch between fused lexical vectors and BERT-generated character embeddings, we first apply a nonlinear projection layer for vector alignment prior to secondary fusion. This dimensional transformation ensures feature space compatibility between the two representation types:

$$\bar{x}_i' = W_2(tanh(W_1 \bar{x}_i + b_1)) + b_2 \tag{3}$$

where, $tanh()$ is the activation function, $z_i$ represents the aligned fused lexical vector, $W_1 \in \mathbb{R}^{d_c * d_w}$ denotes the transformation matrix, $W_2 \in \mathbb{R}^{d_c * d_c}$ is the projection matrix, and $b_1$, $b_2$ are bias terms, where $d_c$ indicates the character embedding dimension and $d_w$ represents the word vector dimension.

Finally, we perform element-wise summation between the aligned fused lexical vectors and the character vectors from BERT's embedding layer, yielding the combined character–word fusion vectors. These integrated vectors $h_i'$ simultaneously encapsulate both character-level and word-level information, as illustrated in Figure 2.

$$h'_i = h_i + \bar{x}'_i \tag{4}$$

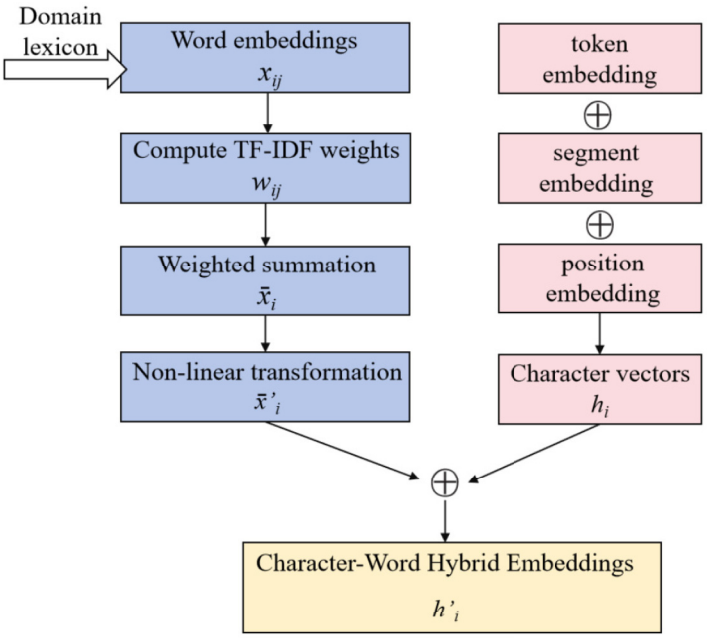



**Figure 2.** Semantic fusion module.

### 3.3. BERT-Enhanced Hybrid Representation

The hybrid character–word embeddings $H = [h'_1, h'_2, \ldots, h'_n]$ are subsequently processed through BERT's Transformer encoder stack. The architecture employs 12-head self-attention mechanisms for dynamic modeling of contextual dependencies, enabling each position to attend to and aggregate the most informative features across the sequence. Specifically, this is achieved by computing three vector representations per token position—query ($Q$), key ($K$), and value ($V$), which interact through scaled dot-product attention:

$$\begin{cases} Q = Linear(H)W_Q \\ K = Linear(H)W_K \\ V = Linear(H)W_V \end{cases} \tag{5}$$

Here, $Linear()$ denotes a linear transformation, and $W_Q$, $W_K$, and $W_V$ represent trainable weight matrices.

The attention computation is formulated as follows:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{6}$$

Here, $Attention()$ denotes the scaled dot-product attention operation. The $\sqrt{d_k}$ term projects the attention matrix into a standard normal distribution space. $Softmax()$ is the normalized exponential function that ensures the attention weights sum to 1.

The model employs multi-head self-attention through parallel computation across multiple attention heads, followed by concatenation of all head outputs and a final linear projection. This architecture facilitates multi-perspective feature extraction from distinct representation subspaces, enables the acquisition of diverse semantic patterns, and ensures comprehensive contextual understanding.

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \tag{7}$$

$$\text{MultiHead}(Q, K, V) = \text{Concat}(head_1, head_2, \ldots, head_k)W^o \tag{8}$$

Let $W_i^Q$, $W_i^K$, and $W_i^V$ denote the query/key/value weight matrices for the $ith$ attention head, with $W^o$ being the output projection matrix and $Concat()$ the concatenation operator.

The sequence $P = [p_1, p_2, \ldots, p_n]$, produced by successive Transformer encoder layers, forms contextualized fusion vectors that holistically encode full-text semantics, making them immediately applicable to named entity recognition tasks.

## 4. ES-BERT Enhanced Named Entity Recognition Model

Our proposed framework for the educational domain NER integrates the enhanced semantic BERT (ES-BERT) model with a BiLSTM-CRF architecture [22–24]. This integration synergistically leverages BERT's deep contextual representations, BiLSTM's capacity for capturing sequential dependencies, and CRF's ability to enforce global label consistency. The model begins with ES-BERT generating semantically enhanced embeddings. These embeddings are then processed by the BiLSTM layer to extract hierarchical contextual features. Finally, the CRF layer decodes the most probabilistically optimal label sequence using the Viterbi algorithm. The entire system is jointly optimized with a focal loss objective to enhance robustness against annotation noise, as illustrated in Figure 3.

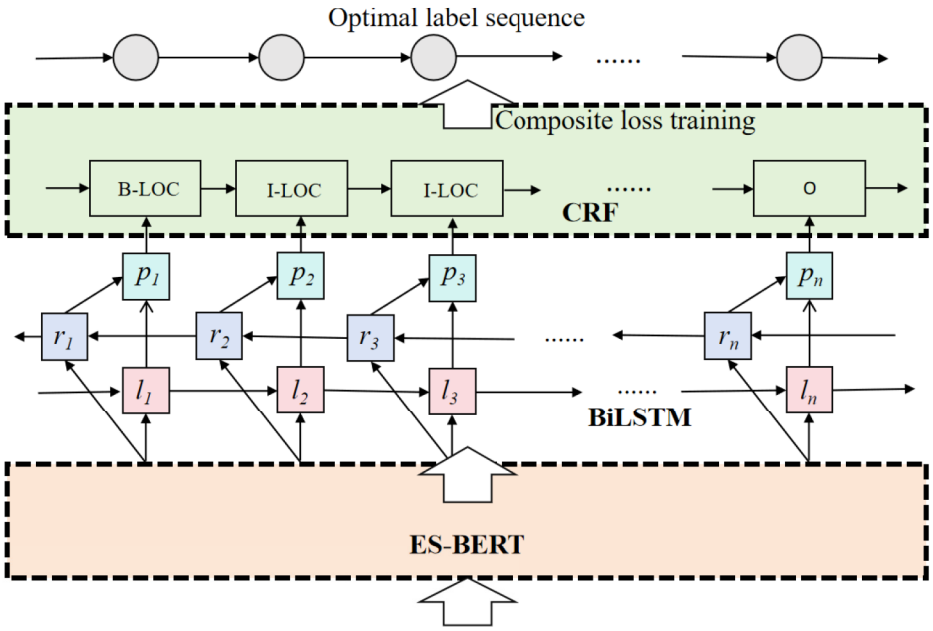



**Figure 3.** Overall structure of the ES-BERT+BiLSTM+CRF model.

### *4.1. BiLSTM Layer*

Standard LSTM networks process sequences unidirectionally (left-to-right), inherently assigning greater importance to later tokens—a problematic bias for segmentation tasks. Our method constructs bidirectional representations through position-wise concatenation of forward $(\overrightarrow{p_1}, \overrightarrow{p_2}, \ldots, \overrightarrow{p_n})$ and backward $(\overleftarrow{p_1}, \overleftarrow{p_2}, \ldots, \overleftarrow{p_n})$ LSTM outputs, formally defined as $L = (l_1, l_2, \ldots, l_n) \in \mathbb{R}^{n*m}$, where each $l_i$ combines both directional contexts for optimized label prediction.

$$\overrightarrow{l_i} = \overrightarrow{LSTM}(p_i) \tag{9}$$

$$\overleftarrow{l_i} = \overleftarrow{LSTM}(p_i) \tag{10}$$

$$l_i = \overrightarrow{l_i} \oplus \overleftarrow{l_i} \tag{11}$$

Here, $\overrightarrow{LSTM}()$ and $\overleftarrow{LSTM}()$ represent the forward and backward LSTM networks, respectively, capturing historical and future contextual information for the current feature position $p_i$. The symbol $\oplus$ denotes vector concatenation. In named entity recognition (NER) tasks, the BiLSTM architecture fully utilizes bidirectional sequence context, significantly improving character-level label prediction accuracy.

### *4.2. CRF Layer*

The CRF layer explicitly models both word-level semantic relationships and inter-label transition probabilities. When the BiLSTM layer generates erroneous predictions, the CRF layer corrects these errors by analyzing neighboring label contexts, thereby refining the final output sequence [25]. Formally, given a label sequence $y = (y_1, y_2, \ldots, y_n)$ and a feature vector $L \in \mathbb{R}^n$, the sequence scoring function is defined as follows:

$$score(y, L) = \sum_{i=1}^{n} E_{y_i} * l_i + \sum_{i=2}^{n} T_{y_{i-1}, y_i} \tag{12}$$

The score at each position is determined by (1) the emission matrix $E$ from BiLSTM outputs, which associates each character with corresponding label weights, and (2) the transition matrix $T$ of the CRF that captures label-to-label movement scores, where $E_i$ denotes the weight vector for label $i$ and $T_{ij}$ indicates the transition score from label $i$ to label $j$.

The Viterbi algorithm's identification of the maximum-probability path is ideally suited for named entity recognition (NER) tasks by globally optimizing sequence likelihood (avoiding local decision pitfalls), enforcing structural constraints through transition probabilities, and capturing long-range dependencies essential for precise entity boundary detection. Formally, given an input sequence $L$ and label space $y$, the optimal tag sequence $\hat{y}$ is computed as follows:

$$\hat{y} = \underset{y'}{argmax} score(y', L) \tag{13}$$

where $y'$ denotes the complete collection of predicted label sequences.

### *4.3. Novel Focal Loss-Based Composite Loss*

The negative log-likelihood loss (NLL), as the standard CRF loss function, demonstrates strong performance under ideal conditions (clean data, balanced class distribution, and complete annotations).

$$L_{CRF} = -log(\frac{exp(score(y, L))}{\sum_{y' \in y^n} exp(score(y', L))}) \tag{14}$$

However, in named entity recognition (NER) tasks, non-entity ("O") labels account for an overwhelmingly high proportion of tokens, while numerous challenging samples (e.g., boundary-ambiguous or nested entities) coexist. When certain entity categories dominate the dataset with significantly larger sample sizes, the negative log-likelihood (NLL) loss introduces substantial gradient bias. This causes the loss function to be predominantly influenced by high-frequency categories, consequently leading to model predictions that are markedly skewed toward majority classes.

Focal loss, originally designed to handle class imbalance and varying classification difficulty in sequence labeling tasks [26], is enhanced in our framework through Viterbi-decoding integration. The proposed method first computes positional alignment discrepancies between Viterbi-derived paths and ground truth annotations, then generates sequence-aware weighting coefficients $\beta_t$. Only samples misclassified under sequential constraints receive elevated weights $\beta_t$. Consequently, we reformulate focal loss as follows:

$$L_{V-FL} = -\frac{1}{T}\sum_{t=1}^{T}\beta_t(1-P_t)^{\gamma}log(P_t) \tag{15}$$

$$s.t. \quad \beta_t = \begin{cases} 1+\alpha & if\ \hat{y}_t \neq y_t \\ 1 & otherwise \end{cases}$$

In the formulation, $T$ represents the sequence length, $p_t$ denotes the probability of the true label at position $t$, and $\gamma$ is the focusing hyperparameter. To address the class imbalance inherent in this task, where majority-class samples are typically classified with ease, a moderate $\gamma$ value (e.g., 2.0) is set to suppress gradient contributions from these simple samples. This focuses the optimization on learning discriminative features, thereby enhancing minority-class performance. The Viterbi weighting factor $\beta_t$ increases when predictions violate sequential constraints, while the hyperparameter $\alpha > 0$ amplifies weights for such misclassified samples. When $\alpha = 0$, the sequence-aware mechanism is inactive; setting $\alpha > 0$ (e.g., 0.75) provides a stronger signal to prioritize correcting errors that disrupt sequential coherence.

During model training, we implement a joint optimization framework that simultaneously preserves the standard CRF negative log-likelihood ($L_{CRF}$) objective and integrates Viterbi-augmented focal loss ($L_{V-FL}$) as an adaptive regularization component (Algorithm 2).

$$L = L_{CRF} + \lambda \cdot L_{V-FL} \tag{16}$$

Here, $\lambda \in [0.1, 1.0]$ is a hyperparameter, with the aim of minimizing the loss.

**Algorithm 2** Viterbi-enhanced focal loss (V-FL)

**Input:** Sequence length $T$;
True label probability $P_t$ at position $t$;
True label sequence $\mathbf{y}_{\text{true}}$;
Viterbi decoding path $\mathbf{y}_{\text{viterbi}}$.
**Output:** Total loss $\mathcal{L}_{\text{total}}$

Compute traditional CRF negative log-likelihood loss $\mathcal{L}_{\text{CRF}}$ from Equation (14);
**Calculate sequence-aware weighting coefficients:**
**for** $t = 1$ to $T$ **do**
  **if** $\mathbf{y}_{\text{viterbi}}[t] \neq \mathbf{y}_{\text{true}}[t]$ **then**
    $\beta_t \leftarrow 1 + \alpha$ *# Increase weight for Viterbi prediction errors*
  **else**
    $\beta_t \leftarrow 1$ *# Base weight for correct predictions*
  **end if**
**end for**
**Compute Viterbi-based focal loss (V-FL):**
$\mathcal{L}_{\text{V-FL}} \leftarrow 0$
**for** $t = 1$ to $T$ **do**
  $\mathcal{L}_{\text{V-FL}} \leftarrow \mathcal{L}_{\text{V-FL}} - \beta_t \times (1 - P_t)^{\gamma} \times \log(P_t)$
**end for**
$\mathcal{L}_{\text{V-FL}} \leftarrow \mathcal{L}_{\text{V-FL}}/T$ *# Normalization*
**Compute joint optimization total loss:** $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CRF}} + \lambda \times \mathcal{L}_{\text{V FL}}$
**return** $\mathcal{L}_{\text{total}}$

This combined approach integrates the sequence-level constraints of the CRF loss with the token-level hard example weighting of the V-Focal loss. The CRF loss inherently models label transition patterns to ensure global coherence, whereas the V-Focal loss emphasizes difficult-to-classify tokens. Their joint optimization achieves two objectives, as follows: (1) globally coherent label sequences and (2) enhanced attention to challenging boundary

cases. As a result, the final predictions exhibit both global sequence consistency and high local classification confidence.

## 5. Experiments

### *5.1. Experimental Environment*

The experimental configuration consisted of an Intel Core i7-8565U processor with 16 GB RAM and 8 GB GPU VRAM, running on Windows 11 with a Python 3.8 runtime environment. The implementation employed PyCharm Professional 2023.2.5 in Anaconda with TensorFlow 2.10.0, and PyTorch 1.13.1, implementing Google's BERT-Base-Chinese pretrained model whose key hyperparameters are detailed in Table 2.

**Table 2.** Model parameter settings.

| Hyperparameters | Values |
|---|---|
| BERT lr | $3 \times 10^{-5}$ |
| BiLSTM lr | $1 \times 10^{-4}$ |
| CRF lr | $1 \times 10^{-3}$ |
| Batch size | 32 |
| Hidden units | 768 |
| Encoder layers | 12 |
| Attention heads | 12 |
| Optimizer | Adam |

### *5.2. Experimental Data*

The experimental dataset comprises a custom-built corpus of educational literature in computer-related disciplines (Edu_literature). We collected 1000 pedagogical research papers from CNKI using web crawlers programmed with topic keywords such as "industry-education integration", "OBE concept", and "project-driven approach". The papers included author affiliations, titles, abstracts, and keywords. Entity annotation was performed through a hybrid approach combining machine processing and manual review. The procedure began with an initial annotation using a pre-trained model-based tool, followed by human-assisted verification to identify and resolve ambiguities and errors. A total of 5312 entities across four categories were annotated. The distribution of these categories is shown in Table 3.

**Table 3.** Categories and quantity of educational literature.

| Entity Types | Count |
|---|---|
| Institution names | 1107 |
| Academic majors | 575 |
| Course titles | 2692 |
| Teaching methods/Pedagogical approaches | 938 |

The dataset employs the standard BIO (begin–inside–outside) annotation scheme, where 'B-X' marks the beginning of entity X, 'I-X' indicates the continuation of entity X, and 'O' denotes non-entity tokens. Annotation examples are shown in Table 4.

**Table 4.** Annotation method of the BIO entity sequence.

| Entity Types | Start | Middle and End | Entity Example |
|---|---|---|---|
| Institution names | B-N | I-N | Peking University, Tsinghua University, Nanjing University, etc. |
| Academic majors | B-M | I-M | Computer Science and Technology, Software Engineering, Digital Media, etc. |
| Course titles | B-C | I-C | Principles of operating systems, Data structures, Digital media technology, etc. |
| Teaching methods/ Pedagogical approaches | B-T | I-T | Project-driven teaching, Integration of industry and education, Flipped classroom, etc. |

### *5.3. Evaluation Metrics*

To assess model performance, we employ three standard evaluation metrics—precision, recall, and F1 score—for named entity recognition analysis.

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

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

$$F1 = \frac{2 \times Pre \times Recall}{Pre + Recall} \tag{19}$$

where TP denotes the number of correctly identified entities, FP represents the number of non-entities incorrectly identified as entities, and FN indicates the number of entities that failed to be recognized.

### *5.4. Entity Recognition and Analysis in Education*

Owing to the limited size of the Edu_literature dataset, a single train–test split risks unstable evaluation due to partitioning variability. To mitigate this, we employed 5-fold cross-validation for all six comparative models. This method ensures that models are trained and validated on nearly the entire dataset in each cycle, maximizing data utility. Performance metrics were averaged over all folds to yield a more robust estimate of model performance and stability.

The average performance metrics across all folds from the 5-fold cross-validation experiments for the models BiLSTM+CRF, BERT+BiLSTM+CRF [9], ALBERT+BiLSTM+CRF [12], Lattice-LSTM [27], Soft-Lexicon (LSTM) [28], and the model proposed in this study are presented in Table 5.

**Table 5.** Experimental results on Edu_literature.

| Model | Pre | Recall | F1 |
|---|---|---|---|
| BiLSTM+CRF | 83.16% | 85.18% | 84.11% |
| BERT+BiLSTM+CRF | 87.11% | 86.37% | 86.71% |
| ALBERT+BiLSTM+CRF | 87.08% | 87.55% | 87.30% |
| Lattice-LSTM | 85.47% | 86.22% | 85.84% |
| Soft-Lexicon (LSTM) | 88.86% | 88.04% | 88.44% |
| Our model | 90.38% | 89.71% | 90.04% |

As shown in Table 5, the model proposed in this study demonstrates superior performance across all three metrics—precision (90.38%), recall (89.71%), and F1 score (90.04%)—significantly outperforming all other comparative models, which reflects its comprehensive advantage. Compared to the baseline model BiLSTM+CRF (F1 = 84.11%), the proposed model achieves an improvement of nearly 6 percentage points in F1 score, indicating the effectiveness of its architectural enhancements. Among the pre-trained language models, ALBERT+BiLSTM+CRF (F1 = 87.30%) slightly surpasses BERT+BiLSTM+CRF (F1 = 86.71%), suggesting that ALBERT's parameter-sharing mechanism may lead to more stable representation learning. However, both fall short of the proposed model, implying that there remains room for optimization beyond simply employing pre-trained models. Within lexicon-enhanced models, Soft-Lexicon (LSTM) (F1 = 88.44%) performs better than Lattice-LSTM (F1 = 85.84%), validating the effectiveness of soft lexical integration. Nevertheless, the proposed model pushes beyond the performance ceiling of such methods, likely due to more refined semantic enhancement mechanisms—such as weighted integration of domain-specific dictionaries—that enable more accurate boundary detection and type recognition. Overall, the proposed model maintains high precision while balancing recall, with its performance advantage likely stemming from targeted modeling of domain-specific characteristics and deep fusion of character- and word-level information, offering an effective new solution for NER tasks.

To scientifically evaluate whether the performance improvement of our proposed ES-BERT model over the classic BERT+BiLSTM+CRF model is statistically significant, we compared the performance of the two models using a paired t-test based on F1 scores from 5-fold cross-validation. The results are presented in Table 6.

**Table 6.** *T*-test: paired two-sample for means.

| Model | M | SD | Correlation Coefficient | t(4) | *p* (Two-Tailed) |
|---|---|---|---|---|---|
| BERT+BiLSTM+CRF | 0.867 | 0.008 | 0.336 | −6.91 | 0.002 |
| Our model | 0.900 | 0.010 | | | |

The paired-sample t-test on the 5-fold cross-validation results (Table 6) revealed that the proposed model (M = 0.900, SD = 0.010) significantly outperformed the BERT+BiLSTM+CRF baseline (M = 0.867, SD = 0.008), t(4) = −6.91, $p = 0.002$. This statistically significant difference indicates that the performance gain afforded by the proposed model is robust and not attributable to random variation. The moderate positive correlation (r = 0.336) between model performances across folds suggests that while both models were influenced by similar data characteristics, the proposed model consistently achieved higher scores. These results provide strong evidence that the proposed model offers a meaningful improvement over the baseline in terms of predictive accuracy.

Based on the comparative experimental results, we further evaluate our model's performance across four entity categories (institutions, majors, courses, and teaching methods/modes) using multiple evaluation metrics, as shown in Figure 4.

Figure 4 demonstrates that our model achieves superior F1 scores for institutions and major entities, while exhibiting relatively lower performance on courses, teaching-method, and teaching-mode categories. In-depth analysis indicates that institutional entities in our self-constructed educational domain dataset exhibit higher frequency and diversity, enabling more comprehensive pattern learning during training. The presence of distinctive lexical markers (e.g., "University"/"College") in institution names further boosts recognition accuracy. The model's high major identification accuracy stems from standardized naming conventions documented in China's Ministry of Education publications, particu-

larly the "Undergraduate Program Catalog" and "Vocational Education Major Catalog", which establish clear naming patterns (e.g., "Computer Science", "Computer Science and Technology"). Course names present greater challenges due to lexical polysemy: terms like "Java Programming", "Web Front-end Development", and "Computer Operating Systems" function ambiguously as both course titles and technical terminology, introducing classification noise that impedes accurate disambiguation. To address these limitations, our future work will focus on the following: (1) dataset expansion and education-domain lexicon enrichment; (2) semantic segmentation precision improvement and named entity boundary detection optimization, ultimately aiming to acquire more precise entity features and boundary information for enhanced recognition performance.

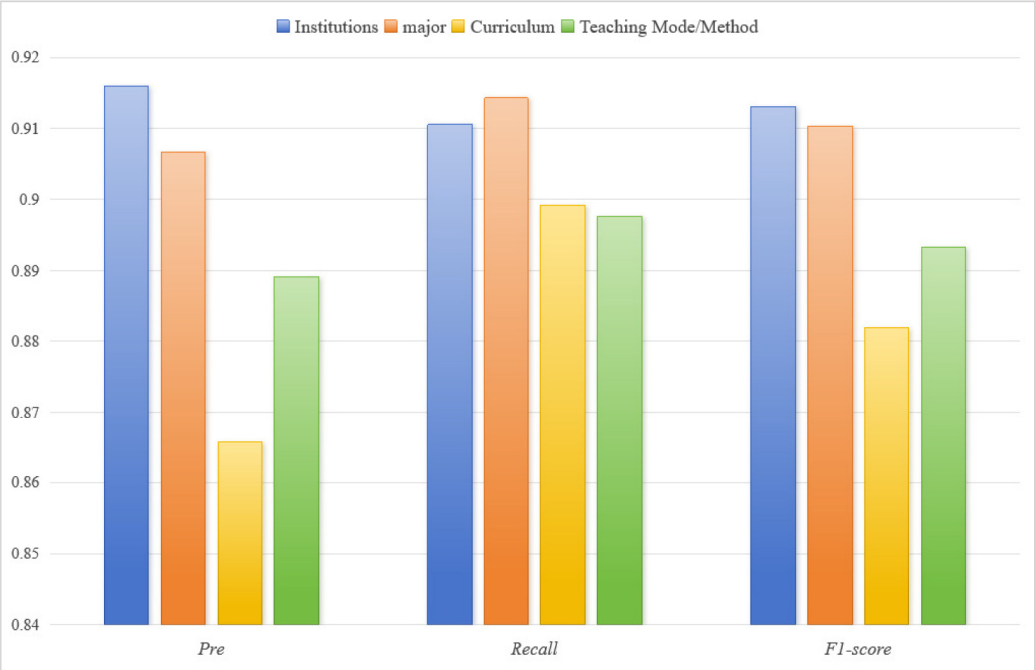



**Figure 4.** Various entity recognition effects.

### *5.5. Ablation Study*

To evaluate the effectiveness of the ES-BERT model with enhanced semantic representation in named entity recognition, four progressively advanced model variants were designed for comparative analysis, with the objective of assessing the contribution of each module:

- Model 1 (Baseline): Uses exclusively BERT-generated character embeddings as input to the BiLSTM-CRF framework, with negative log-likelihood (NLL) loss as the objective function.
- Model 2: Building upon Model 1, semantic enhancement was incorporated without the use of a domain-specific lexicon. A word segmentation tool was applied to generate tokens, which were then fused with character-level representations to form the input to the model.
- Model 3: Based on Model 2, a domain lexicon was integrated along with the proposed ES-BERT model. Domain-relevant vocabulary was combined with character-level information to construct the input representation.
- Model 4 (Full Model): An enhanced version of Model 3, in which a modified focal loss-based joint loss function was introduced to replace the original loss function.

Figure 5 displays the F1 scores obtained from 20 iterative training runs of each model configuration.

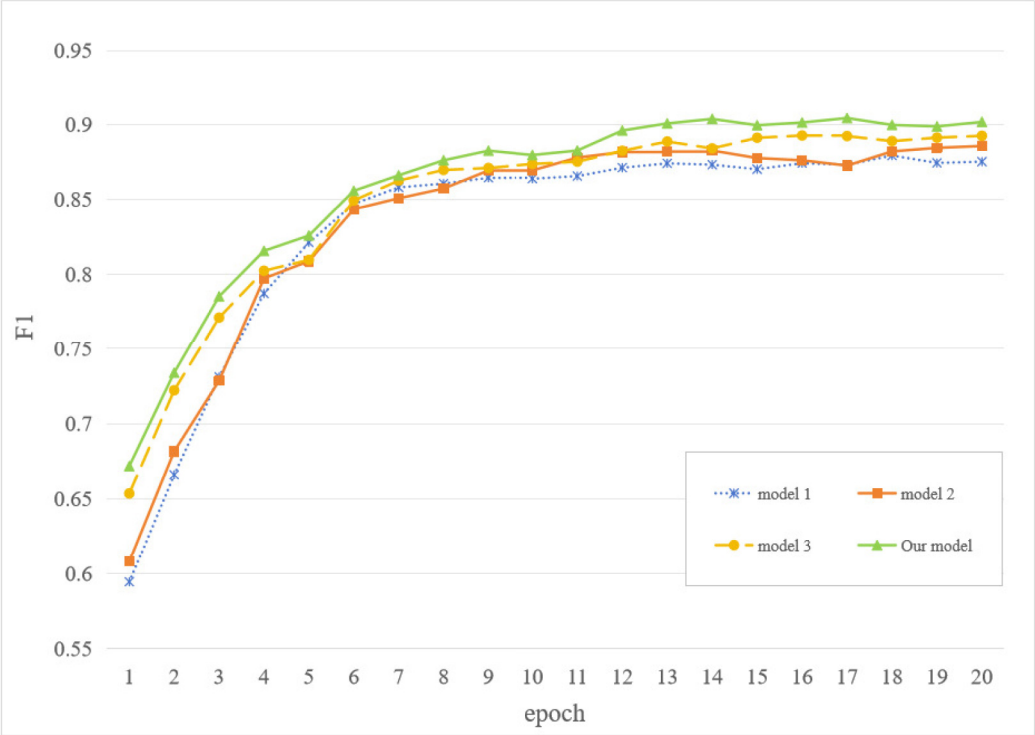



**Figure 5.** The accuracy of ablation experiments.

Figure 5 demonstrates that Model 2 achieves higher overall F1 scores compared to Model 1 after convergence, confirming the effectiveness of lexical information fusion for recognition enhancement. However, our analysis reveals that word segmentation errors cause error propagation, adversely affecting downstream word recognition and, consequently, limiting further performance improvements.

To address this limitation, Model 3 introduces an innovative domain dictionary matching mechanism. By generating character-level candidate word sets with TF-IDF weighting, it significantly reduces the impact of incorrect segmentation—low-weight erroneous segmentations are effectively filtered out, resulting in F1 score improvement over Model 2.

This study introduces an enhanced model based on Model 3, which obtains the optimal F1 score. The model tackles two principal difficulties in domain-specific NER—vague entity boundaries and specialized terms—through strengthened domain semantic encoding, yielding more informative and precise features that augment boundary recognition. Additionally, a modified focal loss function addresses class imbalance by applying increased weights to persistently misclassified examples under contextual constraints, thus refining the learning focus on challenging instances. The collaborative operation of these modules markedly improves recall, consequently elevating the F1 score.

The ablation study confirms the critical importance of component-level synergistic optimization for performance improvement, providing new methodological insights for multi-feature fusion in NER tasks.

### *5.6. Model Transferability*

Comparative experiments were conducted to validate the cross-domain transferability of the proposed model using the open-source KBQA and Resume datasets. Given the lack of dedicated domain-specific lexicons for these datasets, Tencent's open-source Chinese word vectors served as the common domain dictionary for word segmentation in all Trials.

(1) The knowledge base question answering (KBQA) dataset, sourced from the NLPCC-ICCPOL 2016 shared task, comprises a knowledge base with 24,030 entity-attribute

triplets. We split the corresponding QA pairs into training (14,609 pairs), validation (3945 pairs), and test (3945 pairs) sets.

(2) The Resume dataset, contains 16,565 annotated entities across 8 categories (including educational background, locations, and personal names), providing additional scenarios for evaluating transfer learning capability.

Table 7 presents the model's performance on the specialized KBQA (Knowledge Base Question Answering) task set with a triple-based knowledge base, where our model achieves new state-of-the-art benchmarks of 98.41% precision, 98.35% recall, and 98.36% F1 score. These results conclusively evidence the model's Outstanding capability for accurate recognition of diverse entity categories (including persons, organizations, and temporal expressions) in knowledge-intensive QA scenarios. Further analysis indicates the model's dual advantage: maintaining high precision while achieving improved low-frequency entity recognition rates—a breakthrough stemming from the synergistic integration of our ES-BERT architecture with novel focal loss-based joint training optimization. The experimental outcomes not only confirm the architectural superiority of the model but also present a viable technical solution for entity recognition in knowledge base question answering systems.

**Table 7**. Experimental results on KBQA.

| Model | Pre | Recall | F1 |
|---|---|---|---|
| BiLSTM+CRF | 94.54% | 94.44% | 94.49% |
| BERT+BiLSTM+CRF | 96.29% | 96.18% | 96.24% |
| ALBERT+BiLSTM+CRF | 97.94% | 97.90% | 97.92% |
| Lattice-LSTM | 95.52% | 95.31% | 95.41% |
| Soft-Lexicon (LSTM) | 96.90% | 97.02% | 96.96% |
| Our model | 98.41% | 98.35% | 98.36% |

Experimental results on the Chinese Resume dataset (Table 8) confirm the efficacy of the ES-BERT model. By deeply fusing character-level and domain-specific lexical features, the model effectively captures domain knowledge from resumes and precisely identifies entity boundaries (e.g., person names, education, work experience). It achieves state-of-the-art performance (96.75% precision, 96.31% recall, and 96.53% F1 score), demonstrating the validity of our approach for handling the unique linguistic patterns of Chinese resumes. The architecture's adaptive learning mechanism yields highly competitive boundary detection accuracy.

**Table 8.** Experimental results on Resume.

| Model | Pre | Recall | F1 |
|---|---|---|---|
| BiLSTM+CRF | 93.10% | 93.15% | 93.12% |
| BERT+BiLSTM+CRF | 95.06% | 95.15% | 95.10% |
| ALBERT+BiLSTM+CRF | 95.54% | 96.13% | 95.83% |
| Lattice-LSTM | 94.04% | 94.27% | 94.16% |
| Soft-Lexicon (LSTM) | 95.45% | 95.58% | 95.51% |
| Our model | 96.75% | 96.31% | 96.53% |

Systematic evaluation of the cross-domain transfer experiments demonstrates that when employing the general-purpose Tencent open-source Chinese word vectors as a domain lexicon, the proposed model architecture exhibits strong domain adaptation and

generalization capabilities across two distinct datasets. These results establish a solid foundation for the model's application in cross-domain scenarios.

## 6. Conclusions

Advances in AI technology are increasing the importance of named entity recognition (NER) for educational applications—such as adaptive learning, intelligent tutoring, and academic research—establishing it as a critical enabling technology for intelligent education systems. A key limitation in educational NER is insufficient lexical semantic awareness, which leads to weak domain-specific representations and imprecise entity boundaries. To address this, we propose ES-BERT, a novel model that introduces enhanced semantic representation. By integrating a domain-specific semantic enhancement mechanism into BERT's architecture, ES-BERT provides an efficient solution for domain-specific NER.

Nevertheless, the study is subject to certain limitations, particularly the restricted scale of the annotated dataset and the limited coverage of the domain lexicon, underscoring the need for more extensive resources. Several architectural directions warrant further investigation in future work, such as extending lexical-semantic enhancement beyond BERT's lower layers to optimize feature integration strategies, developing dynamic weighting mechanisms for lexical representations across different network depths, and systematically examining the effect of focal loss hyperparameters. Advances in these areas are expected to strengthen the model's ability to comprehend complex semantic structures in the educational domain.

**Author Contributions:** Conceptualization, methodology, software, writing—original draft, writing—review, editing: P.H. and H.Z.; methodology, data curation, writing—original draft, visualization: Y.W. and L.D.; validation, investigation: L.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the following projects: the Research Project on Higher Education Teaching Reform in Jiangsu Province (grant number 2025JGYB405), the Jiangsu Province Industry Education Integration First Class Curriculum Project "Virtual Reality Technology and Application" (grant number 03145031), and the School Level Educational Reform Project of Nanjing University of Science and Technology ZiJin College (grant number 20240103005).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available upon request from the corresponding author: huangping984@njust.edu.cn.

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

## References

1. Zheng, S.; Shen, Y. Educational Named Entity Recognition Integrating Word Information and Self-Attention Mechanism. *Softw. Guide* **2024**, *23*, 105–109.
2. Ren, Y.; Su, B.; Yuan, S. Multidimensional Feature Named Entity Recognition Method in Education Domain. *Comput. Eng.* **2024**, *50*, 110–118.
3. Qiao, B.; Zou, Z.; Huang, Y.; Fang, K.; Zhu, X.; Chen, Y. A joint model for entity and relation extraction based on BERT. *Neural Comput. Appl.* **2022**, *34*, 3471–3481. [CrossRef]
4. Han, Q. Construction of Intelligent Question-Answering System to Improve Knowledge Management Ser-vice from the Perspective of Education Informatization. *J. Inf. Knowl. Manag.* **2025**, *24*, 1–22.
5. Li, Y.; Liang, Y.; Yang, R.; Qiu, J.; Zhang, C.; Zhang, X. CourseKG: An Educational Knowledge Graph Based on Course Information for Precision Teaching. *Appl. Sci.* **2024**, *14*, 2710. [CrossRef]
6. Hu, Y.; Chen, Y.; Huang, R.; Qin, Y. CRF-combined boundary assembly method for biomedical named entity recognition. *Appl. Res. Comput.* **2021**, *38*, 2025–2031.

7. Zhang, R.; Dai, L.; Guo, P.; Wang, B. Chinese Nested Named Entity Recognition Algorithm Based on Segmentation Attention and Boundary-aware. *Comput. Sci.* **2023**, *50*, 213–220.
8. Dong, H.; Kong, Y.; Gao, W.; Liu, J. Named entity recognition for public interest litigation based on a deep contextualized pretraining approach. *Sci. Program.* **2022**, *1*, 7682373.1–7682373.14. [CrossRef]
9. Hu, H.; Li, J.; Dong, Z.; Bai, X. Named Entity Recognition Method in Educational Technology Field Based on BERT. *Comput. Technol. Dev.* **2022**, *32*, 164–168.
10. Xie, T.; Yang, J.-A.; Liu, H. Chinese entity recognition based on BERT-BiLSTM-CRF mode. *Comput. Syst. Appl.* **2020**, *29*, 48–55.
11. Lu, Z.; Zhao, W.; Yin, G. Named Entity Recognition for Textual Intelligence Based on RoB-ERTa_BiLSTM_CRF. *J. China Acad. Electron. Inf. Technol.* **2024**, *19*, 442–447.
12. Jin, L.; Zhang, Y.; Yuan, Z.; Gao, S.; Gu, M.; Liu, X. Chinese named entity recognition of transformer bushing faults based on ALBERT-BiLSTM-CRF. *IEEE Trans. Ind. Appl.* **2025**, *61*, 2115–2123. [CrossRef]
13. Chen, S.; Dou, Q.; Tang, H.; Jiang, P. Chinese nested named entity recognition based on vocabulary fusion and span detection. *Appl. Res. Comput.* **2023**, *40*, 2382–2386+2392.
14. Liu, W.; Fu, X.; Zhang, Y.; Xiao, W. Lexicon enhanced chinese sequence labeling using BERT adapter. *Int. Jt. Conf. Nat. Lang. Process. Meet. Assoc. Comput. Linguist.* **2021**, *2021*, 5847–5858.
15. Wu, G.; Fan, C.; Tao, G.; He, Y. Entity recognition of electronic medical records based on LEBERT-BCF. *Comput. Era* **2023**, *2*, 92–97.
16. Sheng, L.; Zhang, Y.; Wu, D. Chinese named entity recognition method based on lexical enhancement. *Mod. Electron. Tech.* **2022**, *45*, 157–162.
17. Wang, Y.; Wang, Z.; Yu, H.; Wang, G.; Lei, D. The interactive fusion of characters and lexical information for Chinese named entity recognition. *Artif. Intell. Rev.* **2024**, *57*, 258.1–258.21. [CrossRef]
18. Zhao, J.; Qian, Y.; Wang, K.; Hou, S.; Chen, J. Survey of chinese named entity recognition research. *Comput. Eng. Appl.* **2024**, *60*, 15–27.
19. Yang, P.; Dong, W. Chinese named entity recognition method based on BERT embedding. *Comput. Eng.* **2020**, *46*, 40–45+52.
20. Huang, S.; Sha, Y.; Li, R. A Chinese named entity recognition method for small-scale dataset based on lexicon and unlabeled data. *Multimed. Tools Appl.* **2023**, *82*, 2185–2206. [CrossRef]
21. Chen, S.; Luo, C.; Ouyang, X.; Li, W. A Semantic-enhanced Chinese Named Entity Recognition Algorithm Based on Dynamic Dictionary Matching. *Radio Eng.* **2021**, *51*, 519–525.
22. Zheng, X.; Li, B.; Feng, Z.; Liu, X. Entity recognition of network sensitive words and variants based on BERT-BiLSTM-CRF. *Comput. Digit. Eng.* **2023**, *51*, 1585–1589.
23. Che, X.; Xu, H.; Pan, M.; Liu, Q.-L. Two-stage learning algorithm for biomedical named entity recognition. *J. Jilin Univ. Eng. Technol. Ed.* **2023**, *8*, 2380–2387.
24. Liu, X.-H.; Xu, R.-Z.; Yang, C.-Y. A Chinese named entity recognition model based on multi-feature fusion embedding. *Comput. Eng. Sci.* **2024**, *46*, 1473–1481.
25. Liao, T.; Gou, Y.; Zhang, S. BERT-BiLSTM-CRF Chinese named entity recognition combined with atten-tion mechanism. *J. Fuyang Norm. Univ. Nat. Sci.* **2021**, *38*, 86–91.
26. Dang, X.; Liu, J.; Dong, X.; Zhu, Z.; Li, F. Named Entity Recognition of Mechanical Equipment Failure for Imbalanced Data. *Comput. Eng.* **2024**, *50*, 104–112.
27. Zhang, Y.; Yang, J. Chinese NER using lattice LSTM. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 15–20 July 2018; pp. 1554–1564.
28. Ma, R.; Peng, M.; Zhang, Q.; Huang, X. Simplify the usage of lexicon in Chinese NER. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 5951–5960.


**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.

electronics

MDPI

*Article*

# A Multi-Level Feature Fusion Network Integrating BERT and TextCNN

**Yangwu Zhang [1,2], Mingxiao Xu [2] and Guohe Li [2,*]**

[1] School of Information Management for Law, China University of Political Science and Law, Beijing 102249, China; yangwuzh@cupk.edu.cn
[2] Department of Computer Science, China University of Petroleum-Beijing at Karamay, Karamay 834000, China; mingxiaox@cupk.edu.cn
* Correspondence: 2023563020@cupk.edu.cn

**Abstract**

With the rapid growth of job-related crimes in developing economies, there is an urgent need for intelligent judicial systems to standardize sentencing practices. This study proposes a Multi-Level Feature Fusion Network (MLFFN) to enhance the accuracy and interpretability of sentencing predictions in job-related crime cases. The model integrates hierarchical legal feature representation, beginning with benchmark judgments (including starting-point penalties and additional penalties) as the foundational input. The frontend of MLFFN employs an attention mechanism to dynamically fuse word-level, segment-level, and position-level embeddings, generating a global feature encoding that captures critical legal relationships. The backend utilizes sliding-window convolutional kernels to extract localized features from the global feature map, preserving nuanced contextual factors that influence sentencing ranges. Trained on a dataset of job-related crime cases, MLFFN achieves a 6%+ performance improvement over the baseline models (BERT-base-Chinese, TextCNN, and ERNIE) in sentencing prediction accuracy. Our findings demonstrate that explicit modeling of legal hierarchies and contextual constraints significantly improves judicial AI systems.

**Keywords:** multi-level; feature fusion; global features; local features; attention; convolution kernels

## 1. Introduction

Job-related crimes refer to criminal offenses committed by public officials within state institutions, state-owned enterprises, or people's organizations. These crimes involve the abuse of state-delegated powers through acts of corruption, fraud, dereliction of duty, infringement of citizens' personal rights, or subversion of state regulations—all punishable under China's Criminal Law. Suppose job-related crimes cannot be dealt with in a timely and fair manner. In that case, it will affect the credibility and efficiency of the government, which is why the anti-corruption campaign aims to crack down on job-related crimes. Although China's rigorous anti-corruption campaigns have yielded significant results in recent years, the adjudication of job-related crimes faces systemic challenges, including protracted trial cycles and severe case backlogs, particularly in grassroots courts. Public judicial data reveals a consistent upward trend in such cases processed by courts and procuratorates. Within traditional judicial workflows, judges manually analyze voluminous historical precedents and legal provisions—a process that is labor-intensive,

time-consuming, and prone to sentencing inconsistencies due to subjective interpretation. These deviations not only undermine judicial fairness but also risk eroding public trust in judicial transparency. Consequently, carrying out research to enhance sentencing efficiency, minimize human interference, and standardize judicial outcomes has become an urgent imperative.

The advances in research on artificial intelligence technology have been rapid in recent years, and they have a significant impact on practical work in the legal field. With the promulgation of the EU Artificial Intelligence Act, Legal Artificial Intelligence has gradually attracted widespread interest from judicial practice and artificial intelligence researchers, gaining rich research paradigms and traditional diversity from other disciplines, such as natural language processing, machine learning, mathematical statistics, and big data science. At the same time, Legal AI has brought efficiency improvements and fair judgments in judicial practices.

Legal AI should play a role in the governance of job-related crimes, especially by fully leveraging the legal semantic capture capabilities of BERT and TextCNN. To address urgent judicial challenges of job-related crimes in China's anti-corruption campaign, this study pursues research objectives, including designing a hierarchical feature fusion network for legal semantics, optimizing domain adaptation for job-related crimes in legal classifications, and advancing AI's role in China's anti-corruption campaign through feature fusion and a hierarchical workflow.

## 2. Related Work

The rapid development of artificial intelligence technology has brought new possibilities to judicial AI [1]. Early systems mostly focused on expert rule libraries or traditional machine learning methods (such as support vector machines), matching historical precedents with structured case features to provide trial references for judges. Goncalves et al. annotated legal texts, including 42 rulings of the European Court of Human Rights [2]. Nallapati et al. proposed that traditional machine learning classifiers have shortcomings in the connection and extraction of relationships based on the argument from unstructured text to structured text [3]. However, such methods rely on manually defined feature engineering, which makes it difficult to deal with the complex semantics and unstructured features of legal texts. Especially when dealing with cases of job-related crimes, suggestions by judicial AI are often rigid due to the variability of the case and the cross-application of legal provisions, resulting in limited practical application effects.

Natural language processing technology, with its advantages in text understanding and semantic analysis, has shown potential in scenarios such as legal Q&A and contract review [4]. Mikolov et al. proposed the word2vec word embedding method to address the inherent shortcomings of traditional word representation methods in word order and idiomatic expression [5]. The CBOW (continuous bag of words) and Skip–Gram models use neural networks to predict words in contextual environments, capture a large number of accurate syntactic and lexical semantic relationships, and learn high-quality distributed representation methods for words, namely, word embedding representation. The introduction of the negative sampling hierarchical softmax method improves the learning speed of dense vectors for word embedding [6]. Aletras et al. combined legal provisions, word embeddings, and parts-of-speech tagging to provide predictions of judgment outcomes by describing the facts of the input legal case [7]. In order to capture more multi-level contextual information, Vaswani et al. proposed an attention mechanism. The Transformer model based on the attention mechanism makes gradient descent more effective and has a particular applicability [8]. BERT (Bidirectional Encoder Representations from Transformers) is composed of multiple layers of Transformer encoders stacked together,

each layer containing a multi-head self-attention mechanism and a feedforward neural network [9]. The self-attention mechanism dynamically calculates the correlation weights between words, enabling the model to focus on global context and local key information simultaneously [10]. BERT uses bidirectional Transformer encoders, where the self-attention mechanism allows each token to directly attend to all other tokens in the input sequence. Every token representation output by BERT in later layers is inherently contextualized, incorporating global and coarse-grained information from the entire sequence to predict crime, thereby capturing long-range dependencies and semantic relationships beyond immediate neighbors. Kim proposed the TextCNN model [11], which uses a convolutional neural network (CNN) model to classify text. The text is transformed into a dense vector representation through word embedding. Then, the discrete vocabulary is mapped to a continuous semantic space, effectively capturing local feature information for the text to be classified [12]. In addition, TextCNN applies multiple convolutional filters of varying widths that slide over the word embedding sequence. Therefore, each filter specializes in detecting specific local patterns (n-grams) within its window size. At the same time, each filter produces a feature map that highlights where its local and fine-grained information is located, enabling quantitative analysis of the case.

In recent years, the LLM (large language model) has developed rapidly, which applies a large amount of text data to train deep learning models that can generate natural language text or understand the meaning of language text [13]. LLMs can fully leverage the advantages of pre-training on large datasets by using computing power and deep learning patterns [14]. Yue proposed the DISC-LawLLM model based on an LLM, which can solve specific tasks in the legal domain [15]. Moreover, Cui proposed using llama and a knowledge graph to construct the ChatLaw model, which has been iterated to the second generation ChatLaw2 MoE [16]. Represented by these models, multiple Chinese legal models have achieved excellent performance in Chinese legal Q&A and case matching [17]. The Lexis+AI platform launched by the Lexis company integrates large language models such as GPT-4, which can generate sentencing recommendations by analyzing variables such as case nature, involved amount, and defendant background [18]. The Law-AI tool developed by the Casetext group uses natural language processing technology to assist lawyers in legal retrieval, following human instructions to work [19]. LLMs have a wide range of application scenarios. They conduct pre-training on massive data through self-supervised learning or semi-supervised learning and then further optimize their performance and ability through fine-tuning and human alignment [20]. The existing legal intelligence applications mainly focus on general legal services scenarios [21], such as case retrieval or legal consultation, and support for sentencing decisions mostly stays at the macro suggestion level.

Expert systems and traditional machine learning methods have some limitations in the process of predicting job-related crimes. Manual rules failed to handle edge cases, e.g., novel corruption tactics, causing higher error rates in complex job-related crimes [22]. Typically, lawyers dedicate months to encoding rules and making systems scalable in response to evolving law [23]. Without contextual knowledge, ambiguous language, such as the terms “gift” and “bribe,” cannot be interpreted, leading to rigid outcomes [24]. Document failures of rule-based systems in common law show inaccuracy in statutory interpretation [25]. Traditional machine learning proved inadequate for job-related crimes, where case variability and contextual semantics, such as abuse of power, demand adaptive reasoning [26]. Compared with expert systems and traditional machine learning methods, word embeddings and convolutional neural networks are more of a focus for researchers. Word embeddings encode words as dense vectors, enabling similarity metrics and distributional semantics [27]. Due to the large size of the dataset, efficient training could

reduce compute costs for the corpora through negative sampling [28]. Learned analogies of word embeddings could help parse legal relationships in job-related crimes, such as man and king or woman and queen [29]. However, due to overfitting in global feature associations, word embeddings are static embeddings [30], i.e., the same vector for all the same word senses, and they fail to encode the windows of words. Therefore, there may be deviations in understanding polysemous words, such as "gift" in legal and everyday contexts. Feature fusion is a method that fully considers both overfitting in global feature associations and the encoding of word windows. Feature fusion in deep learning integrates multiple cues or features representing different aspects of input characteristics to create a more comprehensive feature representation for tasks like classification. The linear combination of feature statistics and its extension to a general nonlinear fusion method involve minimizing a mean-square error and a regularization term [31]. As the combination of features from different layers or branches, it is often implemented via simple operations, such as summation or concatenation. Attentional feature fusion is described by short- and long-bound connections as well as feature fusion within the frontend layer [32].

However, the current research on legal classification prediction lacks an in-depth examination of auxiliary sentencing for the subdivision scenario of job-related crimes, which is characterized by a lack of logical analysis of criminal circumstances, factual determination, and legal relationships. In particular, insufficient attention is paid to research on the feature fusion of criminal circumstances. Job-related crime is a type of criminal offence that arises from the facts and legal consequences of a duty act. Its legal text includes both coarse-grained information for predicting crime and noncrime and fine-grained information for quantitatively analyzing the length of imprisonment of the defendant. This paper proposes a multi-level feature fusion method based on attention and convolutional kernels to address the problem of global and local feature segmentation in legal texts related to job-related crimes. The attention mechanism captures global features by embedding different semantic positions, while the convolutional kernel obtains local features through sliding windows. Multiple granularity features are then fused to enhance the model's ability to perceive semantic information in job-related crime scenarios and improve its predictive performance for job-related crimes.

## 3. Features of Job-Related Crimes

### *3.1. Basis of Sentencing for Job-Related Crimes*

The theory of sentencing for job-related crimes is based on the Criminal Law of the People's Republic of China combined with judicial interpretations, criminal policies, and judicial practice experience to construct a clear hierarchy of punishment systems that balance fairness and efficiency. The sentencing levels for job-related crimes are divided into six categories, i.e., exemption from punishment, 0–1 year, 1–3 years, 3–5 years, 5–10 years, and more than 10 years. Its theoretical basis is rooted in the principle of proportionality between crime and punishment and the criminal trial policy of balancing leniency and severity, aiming to balance the corresponding relationship between social harm, subjective malice, and punishment through scientific and standardized standards.

In specific sentencing practices, the identification of punishment levels is mainly based on the objective harm and subjective malice of criminal behavior. Taking the crime of embezzlement as an example, Article 383 of the Criminal Law clearly shows that the amount of embezzlement between CNY 30,000 and CNY 200,000 is considered "relatively large" and can be sentenced to fixed-term imprisonment of up to 3 years or detention; CNY 200,000 to 3 million is regarded as a "huge amount" and shall be sentenced to imprisonment for not less than 3 years but not more than 10 years; and an amount of over CNY 3 million is considered "exceptionally large" and is subject to imprisonment for more than 10 years

to life imprisonment. This is shown in Table 1. Although this amount standard has rigid requests, judicial interpretations further introduce flexible standards such as "serious circumstances" and "dire circumstances". For example, in cases of embezzlement, disaster relief, poverty alleviation, or adverse social impact, even if the amount does not reach the threshold, it can still be upgraded to punishment, as shown in Table 2.

**Table 1.** Identification of harm caused by embezzlement. .

| Amount Involved (CNY) | Nature of Crime | Punishment |
|---|---|---|
| <5000 | slight | not be registered |
| 5000–30,000 | general | aggravating circumstances |
| 30,000–200,000 | relatively large | less than 3 years |
| 200,000–3,000,000 | huge | between 3 and 10 years |
| >3,000,000 | extremely huge | more than 10 years |

**Table 2.** Aggravating circumstances in job-related crimes.

| Specific Circumstances | Subjective Nature | Objective Harm |
|---|---|---|
| rescue | first offence | transferring illicit money |
| flood prevention | once embezzled | destruction of evidence |
| epidemic control | previously accepted bribes | escaping overseas |
| poverty alleviation | previously misappropriated | sponsoring others to commit crimes |
| illicit money | other criminal records | inability to recover illicit money |

### *3.2. Global and Local Features of Sentencing*

The global features of job-related crimes are the overall information about the legal relationship of job-related crimes obtained after analyzing the entire legal text. This information focuses on the legal relationship regulated by job-related crimes in the text, ignores local information, and has good stability, maintaining appropriate penalties and accountability. The criminal compositions, causal connections, and criminal accountability involved in job-related crimes have legality, specificity, and universality. The subjective purpose is adapted to the consequences of the crime, and the criminal behavior is consistent with the job-related responsibility. The global features of criminal embezzlement are shown in Table 3. The local features of job-related crimes refer to the features obtained in specific areas of the legal text. These can reflect the local structure and detailed information on the conviction and sentencing of job-related crimes. The determination of the amount of embezzlement belongs to local features, and aggravating factors also belong to local features. The combination of these two kinds of features can elevate the level of judgment in job-related crimes. The local features of the criminal embezzlement are shown in Table 4.

In judicial practice, the scientificity and adaptability of sentencing standards should be considered together. On the one hand, the fixed-amount standard may lead to sentencing differences between different areas due to unbalanced economic development. For example, the actual tolerance of "large amounts" in economically developed areas is higher than that in less developed regions, which damages legal unity and cannot demonstrate the power of legal punishment and deterrence. For example, the embezzlement of CNY 30,000 in economically underdeveloped areas affects the poverty relief of hundreds of people, and its criminal consequences are far greater than those in developed regions. On the other hand, the boundary of judges' discretion needs to be further refined through judicial interpretation. The standard stipulates that probation or exemption shall not be applied to job-related crimes involving serious circumstances and destructive social impact, based

on judicial opinions on several issues concerning the strict application of probation and exemption from criminal punishment in handling cases of job-related crimes, thus reducing the discretion space. The dynamic adjustment mechanism is introduced to link sentencing amount with regional per capita income to enhance the standard's adaptability to the times.

**Table 3.** Global features in embezzlement.

| Subjects | Objects | Subjective Factors | Objective Factors |
|---|---|---|---|
| governmental officials | state assets | directly | act of duty |
| trustees for public property | public property | intentionally | embezzlement |
| exploitation of officials | state nature | illegally | fraud |
| non-governmental officials | public nature | not-faulty | concealment |

**Table 4.** Local features in embezzlement.

| Amount Involved (CNY) | Circumstances | Means | Regions |
|---|---|---|---|
| <5000 | rescue | misappropriation | 1-tier cities |
| 5000–30,000 | flood prevention | stealing | 2-tier cities |
| 30,000–200,000 | epidemic control | swindle | 3-tier cities |
| 200,000–3,000,000 | poverty alleviation | rebate | 4-tier cities |
| >3,000,000 | illicit money | indirect | 5-tier cities |

### *3.3. Definition of Classification Task and Sentencing*

According to the sentencing provisions for job-related crimes, they can be clearly classified into six categories, i.e., exemption from punishment, 0–1 year, 1–3 years, 3–5 years, 5–10 years, and more than 10 years. These six categories not only reflect the quantitative grading of the social harm of job-related crimes but also provide clear classification functions for the construction of machine learning models. In judicial practices, judges need to accurately match the text of the job-related crime to corresponding sentence ranges based on the facts of the case. Essentially, this involves mapping unstructured text (such as case descriptions in judgment documents) to preset category labels, which is highly consistent with the framework of multi-classification tasks from input features to output labels.

The choice of multi-classification tasks over regression or binary classification methods is mainly based on the legal requirement of discreteness and the normative needs of judicial decision-making. The law has strict boundaries for sentencing. For example, there is an unbreakable legal difference between "3–5 years" and "5–10 years". Suppose a regression model is used to predict the specific length of a sentence. In that case, it may ignore such structural boundaries due to numerical continuity, resulting in prediction results deviating from legal standards. The regression model may output "4.8 years", but according to criminal law regulations, this result needs to be classified as "3–5 years" instead of being used directly as an independent value. In addition, multi-classification tasks can generate probability distributions for each sentencing interval. For example, the probability of a case being sentenced to "5–10 years" is 75% and "3–5 years" is 20%. This probabilistic output not only provides a reference for judges but also assists them in balancing the rationality of different sentencing options, enhancing the transparency and scientific handling of decision-making.

## 4. Multi-Level Feature Fusion

### *4.1. Attention Mechanism for Global Features*

In the text of job-related crimes, establishing the legal relationship that constitutes job-related crimes affects the adaptation of conviction and sentencing; for example, establishing the correlation between statements such as "the amount of embezzlement is particularly

huge" and "causing national losses" is crucial for predicting the result of sentencing. The pre-trained language model represented by BERT captures deep semantic features related to sentencing in the text through the self-attention mechanism, such as the strong correlation between descriptions such as "multiple instances of embezzlement" and "refusal to return embezzled money" and severe penalty. After encoding the text into a high-dimensional vector, the model maps it to one of six sentencing labels through a fully connected layer, achieving end-to-end learning from text to category of sentencing, as is shown in Figure 1. In the pre-training stage, BERT learns the intrinsic rules of language through a Masked Language Model (LLM) and Next Sentence Prediction (NSP). The Masked LM forces the model to understand contextual logic by randomly masking some vocabulary and predicting the masked words. For example, in Figure 1, "amount" and "country" are masked. NSP enhances the model's understanding of discourse structure by assessing the continuity of sentence pairs. This pre-training endows BERT with powerful semantic representation capabilities, enabling it to extract implicit legal elements from unstructured text.

Establishing the legal relationship of job-related crimes depends on encoding global features, and segment attributes can introduce more global information into the model. In establishing the legal relationship of job-related crimes, there are many aspects of identification, such as subject, object, subjective, and objective factors, for example, "from December 2012 to April 2013, the defendant Xia…". The segment attributes for words are shown in Table 5. The word is marked with a segment attribute, and each segment attribute identifies the structured metainformation in the legal relationship of job-related crimes. The meaning of the segment attribute mark is shown in Table 6.

**Table 5.** Properties of case description fields in job-related crimes.

| Token | Segment Property | Token | Segment Property | Token | Segment Property |
|---|---|---|---|---|---|
| 2012 | CT | 在 | CQ | 以 | CM |
| 年 | CT | 担任 | CQ | 虚开 | CM |
| 12 | CT | 某市 | CQ | 发票 | CM |
| 月 | CT | 某局 | CQ | 方式 | CM |
| 至 | CT | 副局长 | CQ | ， | CM |
| 2013 | CT | 期间 | CQ | 套取 | CF |
| 年 | CT | , | CQ | 公款 | CF |
| 4 | CT | 利用 | CC | 人民币 | CF |
| 月 | CT | 职务 | CC | 3 | CF |
| , | CT | 之 | CC | 万 | CF |
| 被告人 | CQ | 便 | CC | 元 | CF |
| 夏某 | CQ | ， | CC | 。 | CF |

Chinese characters in the corpus have caused non-English content.

**Table 6.** Meanings of segment property tags.

| Segment Property Tags | Meanings | Presenting Text |
|---|---|---|
| CT | time of crimes | digitals representing date |
| CQ | qualification of offences | names and official positions |
| CC | causes of crimes | text of causal relationships |
| CM | means of crimes | words such as "with" prepositions |
| CF | facts of crimes | digitals representing the amount |

The self-attention mechanism can automatically identify key sentencing elements and generate high-dimensional semantic vectors through nonlinear transformations of multi-layer encoders. The Transformer utilizes the marker of CLS representing the global information of the entire legal text, which aggregates vectors of contextual embeddings and

then outputs probability distributions of six labels of sentencing through a fully connected layer. For example, The probability of the model determining that a case belongs to the "5–10 year" category is 80%, providing a reference for judges. During the fine-tuning, by post-training on a specialized dataset for job-related crimes, self-attention can further adapt to the length of legal texts and optimize its ability to distinguish easily confused categories.

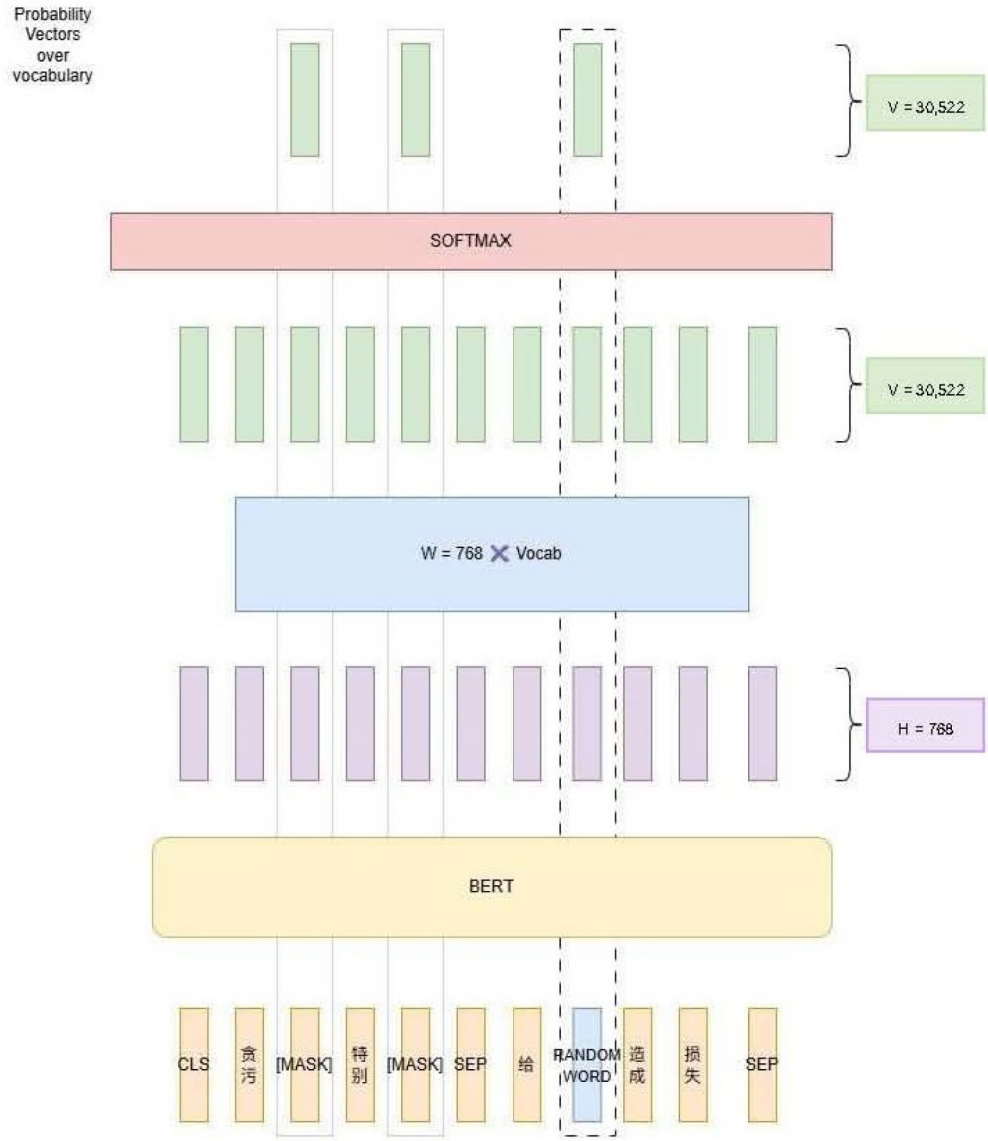



**Figure 1.** Architecture of Bert encoding. Chinese characters in the corpus have caused non-English content.

### *4.2. Convolutional Neural Networks for Local Features*

Local features in the legal text of job-related crimes can be manifested as "after the offences", "illicit funds", "already refunded", etc. Capturing such features can help determine the level of punishment. The architecture of TextCNN mainly includes the embedding layer, convolution layer, pooling layer, and fully connected classification layer, as shown in Figure 2. Firstly, the input text (such as the case description in the judgment document) is transformed into a dense vector representation through word embedding, mapping discrete vocabulary to a continuous semantic space. Subsequently, the convolutional layer uses convolutional kernels of different sizes, performs sliding-window scanning on the text, and then extracts local n-gram features. A $3 \times 1$ convolutional kernel may cover three consecutive words, capturing their combined semantics. The output of each convolutional

kernel is subjected to a nonlinear activation function to generate a feature map. Next, the max-pooling operation is performed to reduce the dimensionality of each feature map, preserving the significant features in each convolutional channel. For example, the feature corresponding to "involved amount of CNY 30,000" is used as the key information of the amount dimension in the legal relationship of job-related crimes. Finally, all pooled features are concatenated into a global vector, which is input into a fully connected layer for classification, which then outputs probability distributions of six labels of sentencing.

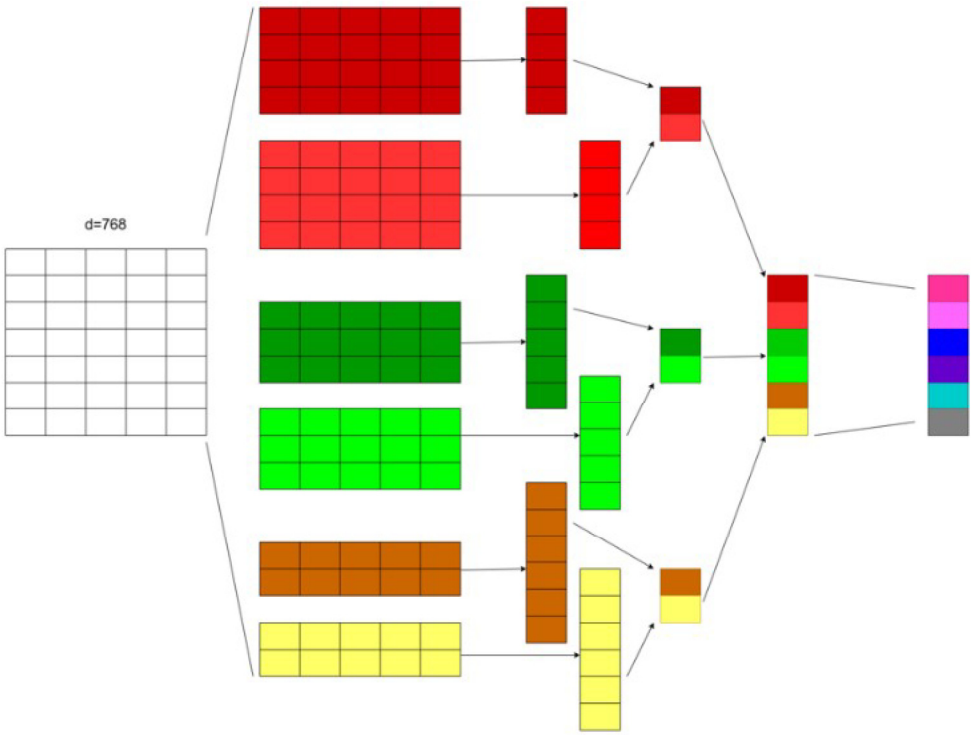



**Figure 2.** TextCNN convolution. Different colors represent different channel information.

### *4.3. MLFFN: Multi-Level Feature Fusion Network*

The Multi-Level Feature Fusion Network needs to consider both the overall structural information of legal relationships in job-related crimes and capture fine-grained local circumstances. Therefore, MLFNN combines the global features on the encoder and the local features on the convolutional kernel by cascading Bert and TextCNN. It could improve the distribution of weighting and responsiveness of feature maps, obtain richer feature representations of job-related crimes, and enhance the model's perception and understanding of conviction and sentencing for job-related crimes. MLFNN is shown in Figure 3.

The applicability of MLFFN in the task of sentencing for job-related crimes is reflected in its ability to map complex legal texts to discrete sentencing categories. Specifically, after the input for case fact description is segmented and embedded (word embedding, segment embedding, position embedding), the Transformer encoders extract semantic features layer by layer, as shown in Formulas (1)–(3), i.e.,

$$ET = \begin{bmatrix} et_{11} & \cdots & et_{1n} & \cdots & et_{1d} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ et_{m1} & \dots & et_{mn} & \dots & et_{md} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ et_{l1} & \cdots & et_{ln} & \cdots & et_{ld} \end{bmatrix}, \quad (1)$$

$$ES = \begin{bmatrix} es_{11} & \cdots & es_{1n} & \cdots & es_{1d} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ es_{m1} & \cdots & es_{mn} & \cdots & es_{md} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ es_{l1} & \cdots & es_{ln} & \cdots & es_{ld} \end{bmatrix}, \tag{2}$$

and

$$EP = \begin{bmatrix} ep_{11} & \cdots & ep_{1n} & \cdots & ep_{1d} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ ep_{m1} & \cdots & ep_{mn} & \cdots & ep_{md} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ ep_{l1} & \cdots & ep_{ln} & \cdots & ep_{ld} \end{bmatrix}. \tag{3}$$

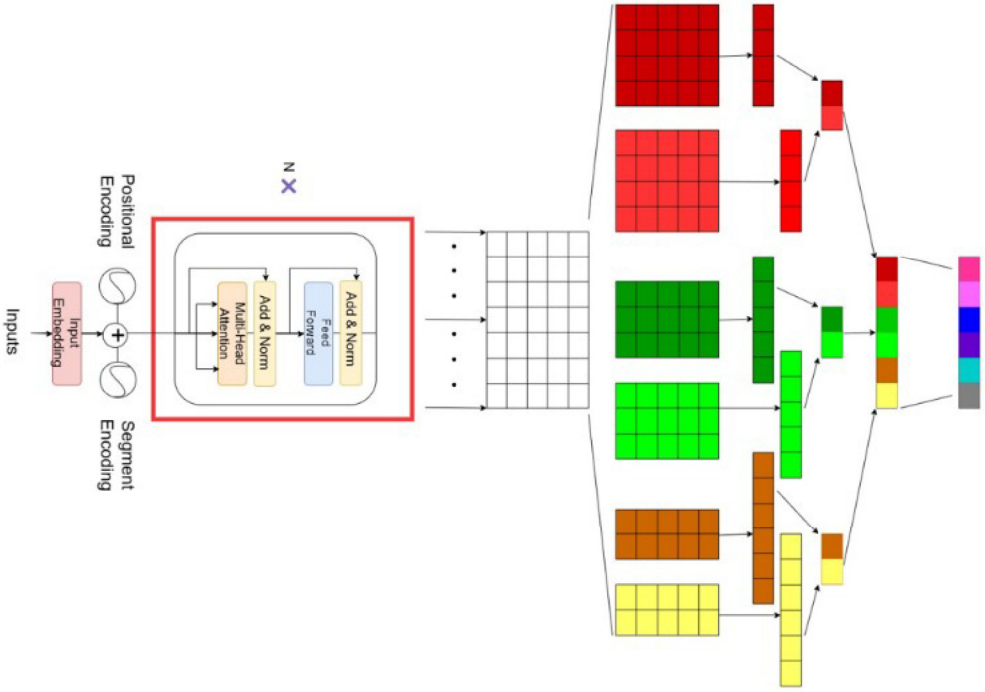



**Figure 3.** MLFNN encoding and convolution. N× represents stacking x identical modules, and the different colors on the right side of the figure indicate different channels.

The hidden vector dimension in the embedding layer is 768. The input word vector can be a GloVe word vector with 100 dimensions, which is mapped into a word embedding ET through a 768-dimensional hidden layer called embeddings of the token, as shown in Formula (1). The input segment vector can be a vector involving the marker of segment property for job-related crimes, and the dimension of the segment vector depends on the legal relationship in the contextual information of job-related crimes. According to Table 6, the dimension of the segment vector with one-hot representation is 5. After 768 dimensions of hidden layer mapping, it is embedded into ES, as shown in Formula (2). The dimension of the input position vector depends on the length of the legal text. Generally, the dimension with one-hot representation is 512, which is mapped to a positional embedding EP through a 768-dimensional hidden layer, as shown in Formula (3). $et_{11}$ represents the first-dimension value of the word embedding for the first token, where $d$ is the dimension of the word embedding. $et_l$ represents the word embedding of the last token in the input sentence, and $et_s$ represents the word embedding of the $m$-th token. The first token can represent

a [CLS] identifier with a special meaning, and the punctuation mark in the middle can be represented as [SEP]. $ep_m$ is the positional embedding of the $m$-th token. $es_m$ is the segment embedding for the contextual information where the $m$-th token is located in the sentence. At the same time, segment embedding can introduce more global information into the model.

The MLFNN frontend encoder can automatically identify key sentencing elements through a self-attention mechanism, using the elements that constitute the legal relationship of job-related crimes as model inputs. After embedding hidden-layer mapping processing, the word embedding ET, segment embedding ES, and position embedding EP are added together, as shown in Formulas (4) and (5), i.e.,

$$EMB = \begin{bmatrix} emb_{11} & \cdots & emb_{1n} & \cdots & emb_{1d} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ emb_{m1} & \dots & emb_{mn} & \dots & emb_{md} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ emb_{l1} & \cdots & emb_{ln} & \cdots & emb_{ld} \end{bmatrix}, \tag{4}$$

and

$$emb_{mn} = et_{mn} + es_{mn} + ep_{mn}. \tag{5}$$

After embedding segment properties that represent the legal relationship of job-related crimes, the MLFFN multi-layer encoders generate high-dimensional semantic vectors containing global feature information through nonlinear transformation. The output encoding corresponding to each token maps word embedding, segment embedding, and position embedding in the opposite end token. Moreover, it includes attention allocation for contextual information, i.e.,

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_i, \dots, head_h)W^o, \tag{6}$$

$$head_i = Attention(Q_i, K_i, V_i), \tag{7}$$

and

$$\begin{aligned} Q_i &= EMB \times W_i^Q, \\ K_i &= EMB \times W_i^K, \\ V_i &= EMB \times W_i^V, \\ Attention(Q_i, K_i, V_i) &= softmax(\frac{Q_i K_i^T}{\sqrt{d_k}})V_i. \end{aligned} \tag{8}$$

It shows that the different attention heads $head_1, \dots, head_i, \dots, head_h$ are concatenated by $Concat(head_1, \dots, head_i, \dots, head_h)$ and transformed into the embeddings input of the next layer by $W^o$ in Formula 6. The three matrices of $Q_i$, $K_i$, and $V_i$ are obtained by linearly transforming the input vector of embeddings, and the corresponding transformation matrices are $W_i^Q$, $W_i^K$, and $W_i^V$, where $W_i^Q$, $W_i^K$, and $W_i^V$ are pre-training parameters of the model, and $W_i^Q \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^K \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^V \in \mathbb{R}^{d_{model} \times d_v}$. The feature map encoded with global information on job-related crimes is output according to Formulas (6)–(8). The MLFNN backend TextCNN uses convolution kernels from sliding windows to filter the global feature map and extract the plot of job-related crimes, making the semantic understanding of local features more accurate. Embedding segment properties that represent the legal relationship of job-related crimes, the MLFFN multi-layer encoders generate high-dimensional semantic vectors containing global feature information through nonlinear transformation. After each token is encoded and output in the MLFNN frontend, it contains segment information on the legal relationship structure. The size of the global feature

map is $l \times d$ (usually $512 \times 768$). The MLFNN backend convolutional layer adopts dual channels with a convolution kernel size of $m \times d$. The output of the channel convolutional layer is calculated according to Formula (9), i.e.,

$$OutConv(i,c,k) = \sum_{p=1}^{r} \sum_{q=1}^{d} w_{p,q} x_{p+i-1,q} \cdot \tag{9}$$

where $w_{p,q}$ represents the connection weight of the convolutional kernel, with a dimension of $r \times d$, and $x_{p+i-1,q}$ represents the value of the $p+i-1$th row of the $qth$ dimension in the global feature map output by the MLFNN frontend. The pooling layer adopts the maximum pooling method and outputs the larger one. Then, the output of the weight by the convolutional kernel is fully concatenated to form a fully connected layer and linked to a softmax layer, which is used for the classification of labels of job-related crimes.

The workflow of the MLFNN model is shown in Figure 4, which includes an input layer, a frontend BERT layer, a backend TextCNN layer, and an output layer. In the input layer, legal documents like the indictment and verdict are split into tokens with position IDs and segment IDs, such as [FACT] and [STATUTE]. Tokenization utilizes WordPiece and legal lexicon expansion, and segment embeddings assign unique IDs to document sections. In the frontend BERT layer, the information of the tokens, the positions, and the segments passes through BERT's Transformer encoders, and multi-head attention captures long-range dependencies to extract global contextual features. The output of the frontend BERT layer is the global encoding sequence, comprising token embeddings, position embeddings, and segment embeddings. In the backend TextCNN layer, parallel convolutions are applied with varying kernel sizes to extract local n-gram patterns from fused features, and max-pooling retains the most salient local features. The output of the backend TextCNN layer concatenates all pooled features. In the output layer, a fully connected layer and softmax activation predict job-related crime categories.

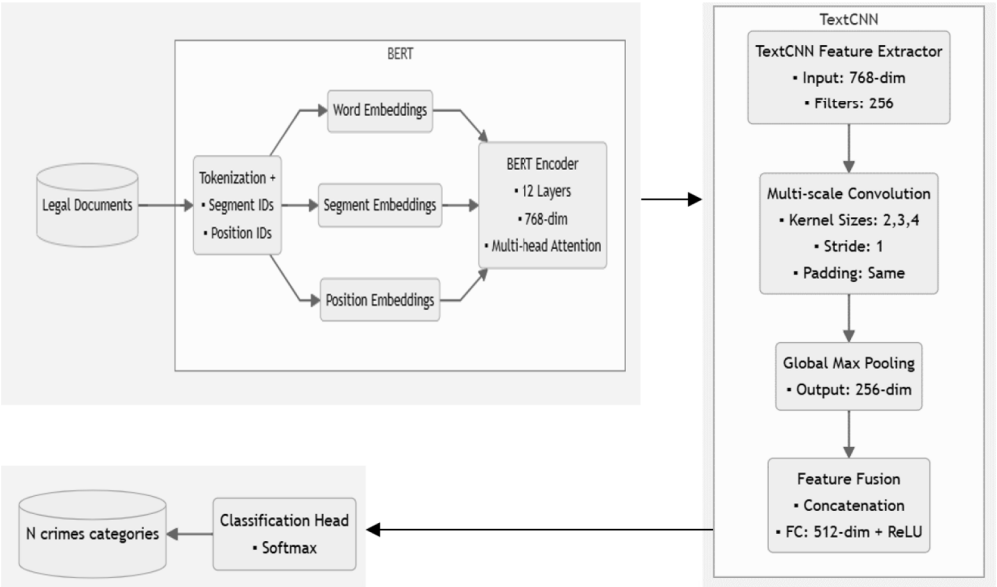



**Figure 4.** Workflow of the MLFNN model.

The MLFNN model differs from other existing models in two aspects. On the one hand, the input of the frontend Bert of MLFNN utilizes tags for words, segments, and positions, corresponding to word embeddings, segment embeddings, and position embeddings, respectively. The segment embeddings are related to criminal circumstances, as indicated in Tables 5 and 6, according to the encoding. The labels for criminal circumstances differ from those in general knowledge and are designated according to legal provisions. Therefore,

the output vectors of the hidden layer of the MLFNN frontend statistically include logical relationships between legal entities. On the other hand, the frontend of MLFNN freezes model parameters except for the last hidden layer, which can be fine-tuned during training based on the crime dataset to provide more contextual information for the MLFNN backend TextCNN to capture important information for legal application classification. The MLFNN algorithm is shown in Algorithm 1.

**Algorithm 1** MLFNN algorithm.

**Require:** bert_model: pre-trained bert model; textcnn_model: input_dim is equal to bert_output_dim; loss_fn: criteria of training with cross entropy loss;
**Ensure:** optimizer: optimal state or approximate optimal state;

```
bert_model = AutoModel.from_pretrained("bert-base-chinese")
# frozen bert parameters
for param in bert_model.parameters() do
    param.requires_grad = False
end for
inputs=tokenizer(texts,token,segment,position)
outputs=bert_model(inputs,output_hidden_states=True)
dataset=load_dataset(outputs)
optimizer=optim.Adam(textcnn_model.parameters(),lr=lr)
criterion=nn.CrossEntropyLoss()
for epoch in range(epochs) do
    textcnn.train()
    total_loss=0
    for i in range(0,len(train_texts),batch_size) do
        embeddings=bert_model(batch_texts)
        logits=textcnn_model(embeddings)
        loss=criterion(logits,batch_labels)
        optimizer.zero_grad()
        loss.backward()
        optimizer.step()
        total_loss+=loss.item()
    end for
end for
```

## 5. Experiments and Results

### *5.1. Data Description*

The current datasets of legal cases mostly come from China Judgments Online [33], the Bank of the National Judicial Examination Center [34], and Challenge of AI in Law [35]. China Judgments Online covers various types of cases, including civil, criminal, administrative, etc., involving factual determination, evidence and basis for the determination, and verdict results. The National Judicial Examination Center provides teaching cases and standardized question banks selected by legal scholars. The case dataset from China Judgments Online and National Judicial Examination Center needed to be obtained through web crawler tools. Unlike the two dataset sources mentioned above, Challenge of AI offers an authoritative dataset of the judicial field that can be publicly downloaded. Therefore, the datasets of job-related crimes come from cases from China Judgments Online, the question bank from the National Judicial Examination Center, and the CAIL2018 (Challenge of AI in Law) dataset.

Data cleaning and data preprocessing are important and interconnected steps in preparing data for experiments. Personal privacy in cases involving sensitive data should not be exposed to the public; therefore, the content of cases related to personal privacy has been desensitized through technological means. The facts of the cases and sentencing

standards, after being cleaned and preprocessed, comply with the sentencing circumstances stipulated in the Criminal Law. The structured data ensures a one-to-one correspondence between criminal facts and sentencing recommendations, providing natural attributes and labels for the corresponding relationship between input and output during modeling. Each case only contains two fields, namely, "statement of fact" and "range of verdict and sentence". The ranges of verdicts and sentences are strictly divided into six categories according to the criminal law, namely, exemption from punishment, 0–1 year, 1–3 years, 3–5 years, 5–10 years, and more than 10 years.

The datasets of job-related crimes involved in this paper are specialized resources containing approximately 270,921 Chinese legal documents related to job-related crimes. The cases used for the experiments are divided into a training set, a verification set, and a test set, with a ratio of 6:2:2. This includes 162,552 cases in the training dataset, 54,185 cases in the verification dataset, and 54,184 cases in the test dataset.

### *5.2. Evaluation Metrics*

On the one hand, legal judgments require both precision and reliability. On the other hand, legal predictions rely on extensive historical case data, and artificial intelligence is revolutionizing the legal profession by enabling attorneys to predict case outcomes with greater accuracy. Although accuracy reflects the overall correctness of the prediction, it is sensitive to data with imbalanced categories, as shown in Formula (10). TP represents a true positive, where the model predicts a positive sample that is actually positive. TN represents a true negative, where the model predicts a negative sample that is actually negative. FP represents a false positive, where the model predicts a positive sample that is actually negative. FN represents a false negative, where the model predicts a negative sample that is actually positive. To objectively describe the model's performance, precision, recall, and F1 score are used as evaluation metrics, as shown in Formulas (11)–(13). Precision emphasizes the reliability of predicting positive classes to avoid wrongful judgments, suitable for scenarios that require reducing false positives. Recall ensures that sufficient positive class samples are identified to avoid missed judgments. The F1 value is the harmony between precision and recall, meeting the rigid demand for fairness in judicial scenarios.

The experimental results are related to the multi-classification task of predicting the sentence of related job crimes, and the dataset has a class imbalance. For example, the number of cases with judgments more than 10 years reached 36,653 in the test set, accounting for 70.2%. Therefore, weighted average precision (WAP), weighted average recall (WAR), and weighted average F1 value (WAF1) are used as performance evaluations, as shown in Formulas (14)–(16). Additionally, $P_i$ and $R_i$ indicate the precision and recall of label $i$, respectively. $count(label_i)$ and $count(dataset)$ represent the count of label $i$ and the count of the dataset in Formulas (14), (15), and (16), respectively.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{10}$$

$$Precision = \frac{TP}{TP + FP}, \tag{11}$$

$$Recall = \frac{TP}{TP + FN}, \tag{12}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}, \tag{13}$$

$$WAP = \sum_{i=1}^{C} \left( \frac{count(label_i)}{count(dataset)} P_i \right), \tag{14}$$

$$WAR = \sum_{i=1}^{C}(\frac{count(label_i)}{count(dataset)}R_i), \tag{15}$$

and

$$WAF1 = \sum_{i=1}^{C}(\frac{count(label_i)}{count(dataset)}F1_i). \tag{16}$$

### *5.3. Experimental Setup*

Bert-base-Chinese, TextCNN, ERNIE, and MLFFN (our model) were selected to conduct experiments evaluating the performance of different models on the dataset of job-related crimes. The dimension of segment embedding of MLFFN frontend is 5, the word embedding adopts a 100-dimensional glove, and the convolution core of the MLFFN backend adopts a dual-channel setting. The experimental environment is provided by AutoDL's cloud service using PyTorch 2.3.0 (https://pytorch.org/get-started/previous-versions/, accessed on 23 August 2025) in the framework and an NVIDIA V100 GPU in the hardware. The hyperparameter setup is shown in Table 7, and the selection of hyperparameters takes into account practical convention, individual experience, and computational power. The dimension of GloVe pre-trained word embeddings used as input to TextCNN is 300, and the convolution kernel size covers two, three, and four windows, with each size configured with 100 filters. The segment embedding dimension of the MLFFN frontend is 5, the positional embedding dimension is 512, the word embedding adopts a 100-dimensional GloVe, and the MLFFN backend convolution kernel adopts a dual-channel setting.

**Table 7.** Setup of Hyperparameters.

| **Models** | **Bert-Base-Chinese** | **TextCNN** | **ERNIE** | **MLFFN** |
|---|---|---|---|---|
| pre-training | Bert-base-Chinese | GloVe | ernie-3.0-base-zh | Bert&GloVe |
| batchsize | 128 | 128 | 128 | 128 |
| training criteria | CE | CE | CE | CE |
| optimizer | AdamW | AdamW | AdamW | AdamW |
| learning rate | $5 \times 10^{-5}$ | $1 \times 10^{-3}$ | $5 \times 10^{-5}$ | $1 \times 10^{-5}$ |

### *5.4. Results and Discussion*

Large models used for comparative analysis have a huge number of parameters and are prone to overfitting training data, remembering noise rather than learning patterns. It is necessary to evaluate the performance of the validation set (such as loss, accuracy, etc.) after each training epoch. Suppose the training set metrics continue to improve while the validation set metrics decrease. In that case, it indicates that the model is beginning to overfit, which requires a timely adjustment of the training strategy.

The training will automatically end when the performance of the validation set no longer improves for more consecutive epochs. At the start of model training, the cross-entropy loss value rapidly decreases. By observing the accuracy curve of models on the validation set, the model reached a stable state after approximately three epochs and 6000 iterations of training, as shown in Figures 5 and 6. The evaluation metrics, including ACC (accuracy), WAP, WAR, and WAF1, of each model on the test set are shown in Tables 8 and 9.

From Table 8, it can be seen that compared to Bert-base-Chinese, MLFFN has a 10%, 6%, and 9% improvement in WAP, WAR, and WAF1, respectively; compared to TextCNN, MLFFN shows improvements of 10%, 14%, and 23% on WAP, WAR, and WAF1, respectively;

and compared to ERNIE, MLFFN has improvements of 6%, 7%, and 10% on WAP, WAR, and WAF1, respectively.

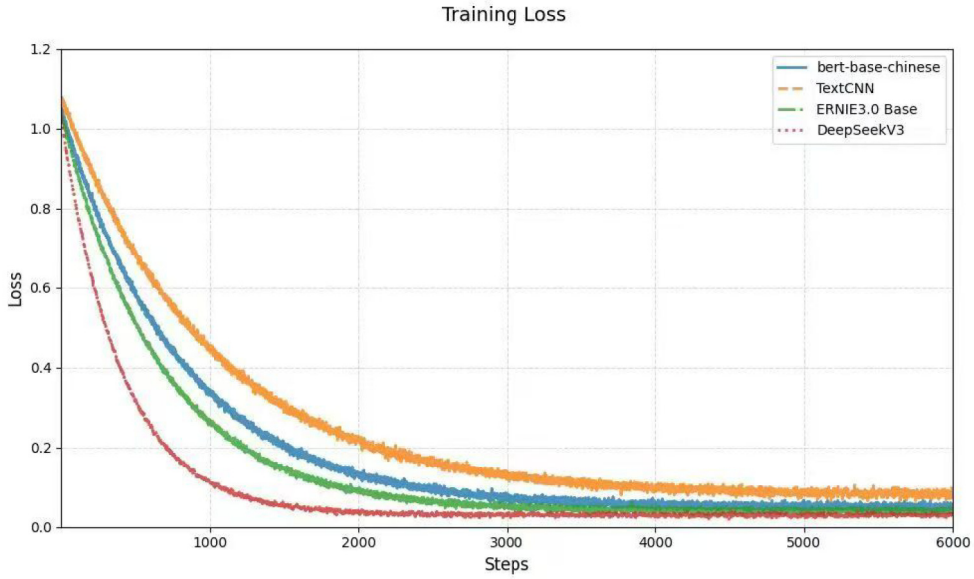



**Figure 5.** Curve of the fine-tuning on the training dataset.

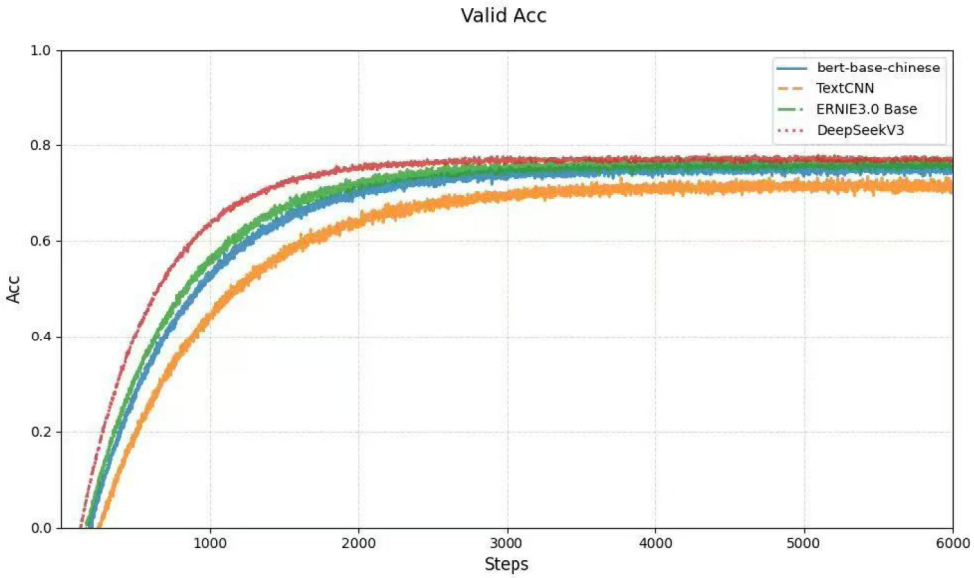



**Figure 6.** Accuracy curve of models on the validation set.

Experimental data shows that the MLFFN model has a performance improvement of over 6% in weighted accuracy, weighted recall, and weighted F1 value, further indicating that the fusion of global and local features is effective in predicting job-related crimes. The model adopted the framework of frontend Bert and backend TextCNN, enhanced with segment embeddings for legal classification in job-related crimes. Key findings include three points as follows.

Firstly, segment embeddings could improve performance by encoding structural boundaries in legal relationships, e.g., sections, clauses, etc. Secondly, the frontend Bert captures long-range contextual semantics in legal relationship clauses of job-related crimes, especially modifiers, while the backend TextCNN extracts discriminative local patterns of job-related crimes, such as legal phrases, particularly those related to duty positions. Thirdly, frozen parameters in the frontend Bert and fine-tuning parameters in the backend

TextCNN could help converge training loss and validation accuracy in the training and validation sets.

**Table 8.** Experimental result of the fine-tuning experiments.

| Models | ACC | WAP | WAR | WAF1 |
|---|---|---|---|---|
| Bert-base-Chinese | 0.7654 | 0.7317 | 0.7654 | 0.7177 |
| TextCNN | 0.7132 | 0.7293 | 0.7132 | 0.6344 |
| ERNIE | 0.7625 | 0.7502 | 0.7625 | 0.7144 |
| MLFFN (ours) | 0.7683 | **0.7525** | **0.7683** | **0.7300** |

**Table 9.** Performances of Bert-base-Chinese, TextCNN, ERNIE, and MLFFN.

| Models | ACC | WAP | WAR | WAF1 |
|---|---|---|---|---|
| Bert-base-Chinese | 0.7654 | 0.7317 | 0.7654 | 0.7177 |
| TextCNN | 0.7231 | 0.7293 | 0.7132 | 0.6344 |
| ERNIE | 0.7152 | 0.7502 | 0.7625 | 0.7144 |
| MLFFN (ours) | 0.8156 | **0.8052** | **0.8156** | **0.7854** |

## 6. Conclusions

In this work, a Multi-Level Feature Fusion Network (MLFFN) is proposed, aiming at the problem of consistency of punishment and offenses in job-related crimes. The word vector, segment vector, and position vector are input to the MLFFN frontend, and the global feature encoding that can represent the legal relationship of job-related crimes is formed through nonlinear mapping of the attention mechanism. Then the MLFFN backend uses convolution kernels with sliding windows to filter the global feature map, and it obtains the local feature information on the special circumstances in job-related crimes. Experiments on the dataset of job-related crimes showed that the MLFFN model has significant performance improvements compared to models such as Bert-base-china, TextCNN, and ERNIE, verifying that the feature fusion method is in line with the applicable rules of benchmark penalties in job-related crimes. In addition, there is some uncertainty in the feature fusion in the post-training of the job-related crimes dataset, which will be the next research direction.

## 7. Limitations and Future Work

Despite MLFFN improving the model's performance, this work has limitations. Segment embedding heavily relies on predefined legal relationships and document structures in job-related crimes, and it may fail to handle them when there are multiple legal relationships. Evaluation focused on the specific dataset, while performance based on other datasets remains unknown. The model utilizes machine learning to learn statistical distributions and patterns from data, replacing legal decision-making logic and reasoning, which, to some extent, poses legal risks.

Future work will be carried out in the following aspects to address these limitations. Distilled BERT variants and dynamic TextCNN filters may be considered to enhance adaptation in other datasets. Learnable segmentation embeddings will be explored for legal relationships in job-related crimes without clear boundaries. Additionally, the framework of Bert and TextCNN will be coupled with attention visualization to provide a rational explanation for predictions.

**Author Contributions:** Conceptualization, G.L.; methodology, G.L.; software, Y.Z. and M.X.; validation, Y.Z. and M.X.; investigation, M.X.; resources, M.X.; data curation, M.X.; writing—original draft

preparation, G.L.; writing—review and editing, Y.Z.; visualization, M.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Research Foundation of China University of Petroleum-Beijing at Karamay (NO. XQZX20240032) and the Education Ministry's Collaborative Education Program with Industry (NO. 231007632264016).

**Data Availability Statement:** The dataset used in this paper is available from the corresponding author.

**Acknowledgments:** We are greatly indebted to colleagues at the Haii (Human–AI Interaction laboratory), Department of Computer Science, Durham University, UK. We thank Zhaoxing Li, Yunzhan Zhou, Chenghao Xiao, Ziqi Pan, Jindi Wang, and Jialin Yu for their special suggestions and many interesting discussions. Additionally, we would like to thank Zengwei Gao, Minjie Zhang, Zhiyang Jia, and Haitao Shi from the Department of Computer Science at China University of Petroleum-Beijing at Karamay for their valuable contribution to this research.

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

## References

1. Arshad, H.; Abdullah, S.; Alawida, M.; Alabdulatif, A.; Abiodun, O.I.; Riaz, O. A Multi-Layer Semantic Approach for Digital Forensics Automation for Online Social Networks. *Sensors* **2022**, *22*, 1115. [CrossRef] [PubMed]
2. Goncalves, T.; Quaresma P. Evaluating preprocessing techniques in a text classification problem. In Proceedings of the Conference of the Brazilian Computer Society, Rio de Janeiro, Brazil, 24–27 October 2005; pp. 525–536.
3. Nallapati, R.; Manning C.D. Legal docket classification: Where machine learning stumbles. In Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, Honolulu, HI, USA, 25–27 October 2008; pp. 438–446.
4. Ali Qutieshat, E.M.; Ali Quteishat, A.M.; Qtaishat, A. Transforming the Judicial System: The Impact of Machine Learning on Legal Processes and Outcomes. *Int. J. Relig.* **2024**, *5*, 6833–6841. [CrossRef]
5. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.; Dean, J. Distributed representations of words and phrases and their compositionality. In Proceedings of the 26th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–10 December 2013; pp. 3111–3119.
6. Mikolov, T.; Chen, K.; Corrado G.; Dean, J. Efficient estimation of word representations in vector space. In Proceedings of the 2013 International Conference on Learning Representations, Scottsdale, AZ, USA, 2–4 May 2013; pp. 525–533.
7. Aletras, N.; Tsarapatsanis, D.; Preoiuc-Pietro, D.; Lampos, V. Predicting judicial decisions of the European Court of Human Rights: A Natural Language Processing perspective. *PeerJ Comput. Sci.* **2016**, *2*, e93. [CrossRef]
8. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 6000–6010.
9. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186.
10. Christoph, F.; Axel, P.; Andrew, Z. Convolutional Two-Stream Network Fusion for Video Action Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June 2016; pp. 933–1941.
11. Kim, Y. Convolutional Neural Networks for Sentence Classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 25–29 October 2014; pp. 1746–1751.
12. Yu, C.Q.; Wang, J.B.; Peng, C.; Gao, C.; Yu, G.; Sang, N. BiSeNet: Bilateral Segmentation Network for Real-Time Semantic Segmentation. In Proceedings of the 15th European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 334–349.
13. Kojima, T.; Gu, S.; Reid, M.; Matsuo, Y.; Iwasawa, Y. Large Language Models are Zero-Shot Reasoners. In Proceedings of the 36th Conference on Neural Information Processing Systems, New Orleans, LA, USA, 28 November–9 December 2022; pp. 936–943.
14. Yang, A.; Xiao, B.; Wang, B.; Zhang, B.; Bian, C.; Yin, C.; Lv, C.; Pan, D.; Wang, D.; Yan, D.; et al. Baichuan 2: Open large-scale language models. *arXiv* **2023**, arXiv:2309.10305. [CrossRef]
15. Yue, S.; Liu, S.; Zhou, Y.; Shen, C.; Wang, S.; Xiao, Y.; Li, B.; Song, Y.; Shen, X.; Chen, W.; et al. LawLLM: Intelligent legal system with legal reasoning and verifiable retrieval. In Proceedings of the 29th International Conference on Database Systems for Advanced Applications, Gifu, Japan, 2–5 July 2024; pp. 304–321.
16. Cui, J.; Ning, M.; Li, Z.; Chen, B.; Yan, Y.; Li, H.; Ling, B.; Tian, Y.; Yuan, L. ChatLaw: A multi-agent collaborative legal assistant with knowledge graph enhanced mixture-of-experts large language model. *arXiv* **2024**, arXiv:2306.16092.

17. Touvron, H.; Lavril, T.; Izacard, G.; Martinet, X.; Lachaux, M.-A.; Lacroix, T.; Rozi è re, B.; Goyal, N.; Hambro, E.; Azhar, F.; et al. LLaMA: Open and efficient foundation language models. *arXiv* **2023**, arXiv:2302.13971. [CrossRef]
18. Lexis. Available online: https://www.lexisnexis.com/pdf/lexis-plus-ai-user-guide.pdf (accessed on 1 June 2024).
19. Create AI. Available online: https://creati.ai/ai-tools/contract-lifecycle-management/ (accessed on 3 May 2024).
20. Cobbe, J. Administrative law and the machines of government: Judicial review of automated public-sector decision-making. *Leg. Stud.* **2019**, *18*, 890–905. [CrossRef]
21. Ekeh, A.; Apeh, C.E.; Odionu, C.; Austin, B. Automating Legal Compliance and Contract Management: Advances in Data Analytics for Risk Assessment, Regulatory Adherence, and Negotiation Optimization. *Eng. Technol. J.* **2025**, *10*, 3684–3703. [CrossRef]
22. Niklas, B.; Nicolas, G.; Frick R.A. Identifying and Generating Edge Cases. In Proceedings of the 2nd ACM Workshop on Secure and Trustworthy Deep Learning Systems, Singapore, 2–20 July 2024; pp. 16–23.
23. Witt, A.; Huggins, A.; Governatori, G.; Buckley, J. Encoding Legislation: A Methodology for Enhancing Technical Validation, Legal Alignment and Interdisciplinarity. *Artif. Intell. Law* **2024**, *32*, 293–324. [CrossRef]
24. Surden, H. Values Embedded in Legal Artificial Intelligence. *IEEE Technol. Soc. Mag.* **2022**, *41*, 66–74. [CrossRef]
25. De Luca, E.W.; Fiorelli, M.; Picca, D.; Stellato, A.; Wehnert, S. Legal Information Retrieval meets Artificial Intelligence (LIRAI). In Proceedings of the 34th ACM Conference on Hypertext and Social Media, Rome, Italy, 4–8 September 2023; pp. 461–464.
26. Zhang, C.; Meng, Y. Bayesian Deep Learning: An Enhanced AI Framework for Legal Reasoning Alignment. *Comput. Law Secur. Rev.* **2024**, *55*, 109267–109279. [CrossRef]
27. Ustun, A.; Can, B. Incorporating Word Embeddings in Unsupervised Morphological Segmentation. *Nat. Lang. Eng.* **2021**, *27*, 609–629. [CrossRef]
28. Lindgren, E.M.; Reddi, S.; Guo, R.; Kumar, S. Efficient Training of Retrieval Models Using Negative Cache. In Proceedings of the 35th Annual Conference on Neural Information Processing Systems, Electr Network, Online, 6–14 December 2021; pp. 1–13.
29. Allen, C.; Hospedales, T. Analogies Explained: Towards Understanding Word Embeddings. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 223–231.
30. Sasaki, S.; Heinzerling, B.; Suzuki, J.; Inui, K. Examining the Effect of Whitening on Static and Contextualized Word Embeddings. *Inf. Process. Manag.* **2023**, *60*, 103272–103289. [CrossRef]
31. Salazar, A.; Safont, G.; Vergara, L.; Vidal, E. Graph Regularization Methods in Soft Detector Fusion. *Access* **2023**, *11*, 144747–144759. [CrossRef]
32. Dai, Y.; Gieseke, F.; Oehmcke, S.; Wu, Y.; Barnard, K. Attentional Feature Fusion. In Proceedings of the 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–8 January 2021; pp. 3186–3195.
33. Information Center at the Supreme People's Court. Specifications for the Use of Public Data on China Judgements Online. Available online: https://wenshu.court.gov.cn (accessed on 10 April 2025). (In Chineses)
34. National Judicial Examination Center. Exam Classroom. Available online: https://www.moj.gov.cn/pub/sfbgw/jgsz/jgszzsdw/zsdwgjsfkszx/gjsfkskskt/ (accessed on 20 May 2025). (In Chineses)
35. Information Center at the Supreme People's Court. A Large-Scale Legal Dataset for China AI and Law Challenge. Available online: http://cail.cipsc.org.cn (accessed on 20 May 2025). (In Chineses)


**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.

*electronics*

MDPI

*Article*

# Enhancing Neural Collaborative Filtering for Product Recommendation by Integrating Sales Data and User Satisfaction

**Haoyang Xia [1,2,*] and Yuanyuan Wang [1,*]**

[1] Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Ube 755-8611, Japan
[2] MICWARE Co., Ltd., Kobe 650-0035, Japan
* Correspondence: f003wcw@yamaguchi-u.ac.jp (H.X.); y.wang@yamaguchi-u.ac.jp (Y.W.)

**Abstract**

The rapid growth of e-commerce has made it increasingly difficult for users to select appropriate products due to the overwhelming amount of available information. Although many platforms, such as Amazon and Rakuten, encourage users to leave reviews, effectively utilizing this information for personalized recommendations remains a challenge. To address this issue, we propose a multi-task product recommender system that supports both new users without purchase histories and existing users with interaction records. For new users without purchase histories, we introduce a ranking-based method that combines three market-oriented features: sales volume, sales period, and user satisfaction. User satisfaction is quantified using sentiment analysis of product reviews. These three factors are integrated into a composite score to identify products with a strong market presence and positive customer feedback. For existing users, we developed an enhanced neural collaborative filtering (NCF) method that incorporates a product bias factor. This model, named bias neural collaborative filtering (BNCF), utilizes multilayer perceptrons to learn latent user–product interactions while also capturing item popularity bias. We evaluated the proposed approaches using a real-world dataset from Rakuten. The results show that our multi-task system effectively improves recommendation quality for users in both cold-start and data-rich scenarios.

**Keywords:** neural collaborative filtering; bias-aware recommendation; cold-start user recommendation; personalized product ranking; sentiment-aware recommendation; Bayesian personalized ranking; deep learning for recommender systems

## 1. Introduction

The rapid growth of e-commerce has significantly altered consumer behavior by providing the convenience of a wide selection of products and a personalized shopping experience. A key technology supporting this transformation is the recommender system, which helps users to discover relevant products by analyzing their preferences, behavior, and contextual data. Product reviews and ratings are valuable sources of user information because they contribute to product credibility by providing subjective feedback. Lackermair et al. [1] demonstrated that online reviews and ratings directly impact purchase intentions, establishing them as a primary source of consumer decision-making. However, due to the volume and unstructured nature of this feedback, users often struggle to process it efficiently. Changchit et al. [2] further emphasized that integrating customer reviews with product specifications enhances customer satisfaction and influences purchasing decisions.

Although review data is valuable, it is insufficient to solve all recommendation challenges, especially those involving new users with limited historical interactions. To address this cold-start problem, we propose a recommendation approach that integrates three market-oriented product attributes: sales volume, sales period, and user satisfaction. The user satisfaction score is calculated using sentiment analysis of product reviews and rating distributions. These attributes are combined using a weighted ranking model with threshold-based filtering. It enables the system to recommend products that are both popular and have positive customer reviews.

Many recommender systems, particularly those on e-commerce platforms, are based on collaborative filtering (CF). However, the effectiveness of CF is often influenced by user search and browsing behavior. Illm et al. [3] pointed out that CF performance degrades when users explore broad product categories or lack specific interests. To address data sparsity and improve accuracy, researchers have extensively studied hybrid approaches that combine CF with content-based filtering. Çano et al. [4] demonstrated that these combinations effectively mitigate the cold-start problem and enhance personalization. Additionally, Sarwar et al. [5] showed that item-based CF methods outperform user-based methods in terms of scalability and robustness.

Recent advancements in deep learning have further expanded the capabilities of recommender systems. Deep neural networks (DNNs), which are commonly used for image recognition and natural language processing, can now model user–item interactions. These models can capture complex nonlinear relationships and learn from implicit feedback data. According to Ko et al. [6], the use of deep learning has improved recommender performance by addressing issues such as data sparsity and hidden user preferences. However, challenges remain, particularly regarding model scalability and interpretability. As consumer behavior becomes more dynamic and context-dependent, recommendation algorithms must incorporate both implicit behavioral signals and explicit market indicators.

In this study, we propose a dual-framework recommender system that addresses both cold-start and data-rich scenarios. For new users without purchase histories, we propose a ranking-based approach for recommending products that considers product attributes, such as sales volume, sales period, and user satisfaction. For existing users with historical interactions, we develop a bias neural collaborative filtering (BNCF) model that incorporates item-level popularity factors into the recommendation process. The detailed workflow and model architecture are described in Section 3.

In summary, our contributions are twofold:

- We propose a ranking-based recommendation approach for new users that integrates sentiment analysis with sales-oriented features. This approach helps to overcome the cold-start problem without relying on user history.
- We introduce a bias-aware neural collaborative filtering model that incorporates product-level popularity factors to improve recommendation performance for users with prior interactions.

The remainder of this paper is structured as follows: Section 2 reviews the related work on traditional and deep learning–based recommender systems, emphasizing sentiment analysis and hybrid models. Section 3 describes the proposed recommendation approaches for both new and existing users, including feature extraction and model design. Section 4 presents the experimental results, including performance evaluations, ablation studies, and hyperparameter analyses. Section 5 discusses the findings and compares the proposed methods. Finally, Section 6 concludes the paper and outlines directions for future research.

## 2. Related Work

To position our contribution within the broader research landscape, we review the existing studies across three key themes: (1) cold-start and hybrid recommendation strategies, (2) neural collaborative filtering and bias-aware models, and (3) sentiment-aware and multimodal approaches in recommender systems.

### *2.1. Cold-Start and Hybrid Recommendation Methods*

The cold-start recommendation problem has been addressed through content-based filtering, popularity metrics, hybrid methods, and rule-based logic. Traditional collaborative filtering struggles when there is insufficient user history. To solve this issue, researchers have proposed hybrid models that incorporate demographic features, purchase history, and temporal behavior [7,8].

Recent methods integrate additional side information, such as user behavior, contextual cues, and external metadata. For example, Adeniyi et al. [8] employed clickstream data and K-nearest neighbor (KNN) to develop real-time content-based recommender systems. Zou et al. [9] employed formal concept analysis to overcome sparsity. Wu et al. [10] integrated temporal and contextual factors via Bayesian modeling and hierarchical clustering. These approaches demonstrate that combining content, behavior, and structure improves cold-start performance.

Our approach builds on this hybrid logic by combining product-level popularity metrics, such as sales volume and period, with sentiment-derived satisfaction to create a multidimensional ranking score. This method effectively recommends products to new users without requiring interaction data.

### *2.2. Neural Collaborative Filtering and Bias-Aware Enhancements*

Deep learning has revolutionized recommender systems by capturing high-order interactions and nonlinear dependencies. He et al. [11] introduced neural collaborative filtering (NCF), in which the inner product of matrix factorization is replaced with multilayer perceptrons (MLPs). This allows for richer feature representations. Extensions of NCF, such as DeepNCF [12] and Wide & Deep [13], combine generalization and memorization.

Other works enhance traditional collaborative filtering using convolutional neural networks (CNNs) [14] or recurrent architectures (e.g., LSTM) to model temporal dynamics [15]. Bobadilla et al. [7] emphasized the importance of bias modeling in BiasedMF [16], which explicitly incorporates user/item rating deviations.

While these models improve personalization, few directly integrate product-level popularity or bias as structured input. Our BNCF model extends NCF by incorporating item-side attributes (sales volume, period, and satisfaction) as bias signals into the embedding and MLP layers, improving interpretability and cold-start robustness.

### *2.3. Sentiment, NLP, and Multimodal Integration in Recommendations*

Natural language processing (NLP) has become a popular method for extracting insights from user reviews. Redhu et al. [17] and Pham et al. [18] demonstrated that sentiment analysis enhances the performance of recommender systems by capturing emotional and experiential feedback that is not evident in ratings alone. Models such as deep cooperative neural networks (DeepCoNNs) [19] and the review and content-based deep fusion model (RC-DFM) [20] use convolutional neural networks (CNNs) or autoencoders to fuse reviews with product metadata.

Transformer-based models (e.g., BERT) have recently improved sentiment understanding and semantic modeling in recommendation tasks [21,22]. Xu et al. [21] combined BERT with KNN to enhance semantic similarity, while Bellar et al. [22] compared BERT with FastText for emotion-rich predictions.

Our method builds on this foundation by using the Google Cloud Natural Language API for sentiment scoring, which we integrate with user ratings to compute a satisfaction score. This score serves as one dimension of our composite ranking model for cold-start users. Meanwhile, multimodal models such as BoNMF [23] and RC-DFM [20] demonstrate the power of integrating text, image, and behavioral data. Although these systems are impressive, they often require substantial computing power. Our approach integrates rich signals through simpler scalar metrics derived from sentiment, ratings, and market data, maintaining interpretability and efficiency.

## 3. Dual-Approach Recommender System Design

This section outlines our recommendation strategies for two types of users: new users without purchase histories and returning users with purchase records. Different methods are necessary to address the diversity of user behavior on e-commerce platforms and to improve the accuracy and relevance of product recommendations.

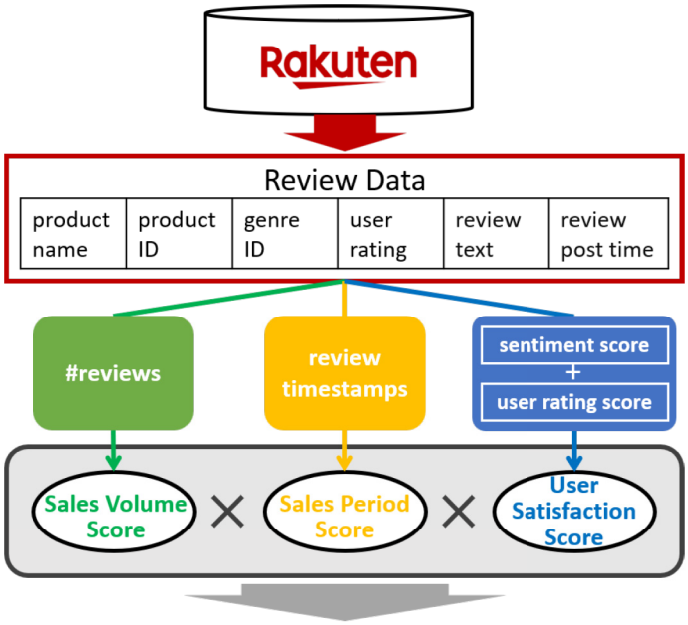
Poriya et al. [24] distinguish between non-personalized and user-based collaborative filtering recommender systems. Non-personalized systems use aggregate metrics, such as product popularity or average ratings, to generate recommendations that are the same for all users. In contrast, user-based collaborative filtering uses algorithms, such as the Pearson correlation coefficient, to analyze similarities between users and recommend products based on the preferences of similar users.

In this study, we propose a dual recommendation strategy. For new users without purchase histories, we utilize a content- and behavior-based approach that integrates product popularity and user feedback. For users with existing purchase data, we apply a neural collaborative filtering model that incorporates user and product bias factors. Our goal is to use natural language processing and deep learning techniques to provide personalized product recommendations to both user types, thereby enhancing the overall shopping experience.

The following subsections describe our approach in detail. Section 3.1 presents the recommendation method for new users. It explains the data selection and preprocessing steps, defines the evaluation metrics of product performance, and introduces the composite recommendation formula. Section 3.2 covers the method for returning users, including data acquisition, constructing the BNCF model, and extracting user- and product-specific features.

### *3.1. Product Recommendation Method for New Users Without Purchase Histories*

This section focuses on recommendation methods for new users who have recently registered on the platform but have not yet made any purchases. In such cases, the absence of purchase behavior and user-specific attributes makes it hard to infer preferences through conventional personalized approaches. Therefore, we propose a method that generates recommendations using objective product performance indicators, such as sales data, user reviews, and market longevity. Figure 1 illustrates the overall workflow of our product ranking system for cold-start users.




| Rank | Product Name |
|---|---|
| 1 | 20-Inch Hardside Spinner |
| 2 | 25-Inch Softside Expandable Roller Luggage |
| 3 | Softside Upright Luggage Set |
| 4 | 28-Inch Hardside Spinner |
| 5 | Luggage Suitcases With Spinner Wheels |
| ⋮ | ⋮ |

**Figure 1.** Workflow of the proposed product ranking method for new users without purchase histories.

In general, products with high sales figures tend to have widespread market acceptance, reflecting qualities such as high quality, good value, and satisfactory performance. Similarly, products that maintain high sales over extended periods demonstrate sustained customer satisfaction and reliability. To capture these aspects, we introduce a recommendation strategy that evaluates products based on three dimensions: sales volume, sales period, and user satisfaction.

We constructed a multidimensional hybrid recommendation model that integrates content-based sentiment analysis with sales and temporal data. This model represents each product in a three-dimensional space, where the axes correspond to the normalized values of sales volume, sales period, and user satisfaction. Products are plotted in this space according to their performance across these attributes. A product that scores highly on all three dimensions occupies a position in the ideal region of this space, visualized in Figure 2 as a red cube. Our recommendation algorithm prioritizes these products, aiming to present users with options that are both popular and well-reviewed over time.

The methodology begins with the collection and preparation of product metadata. When direct sales data is unavailable, we define proxy metrics for sales volume and sales period using the number and timing of user reviews. We also apply sentiment analysis to review texts to quantify user satisfaction. These metrics are then integrated into a unified recommendation score that ranks products according to their overall performance across all three dimensions.

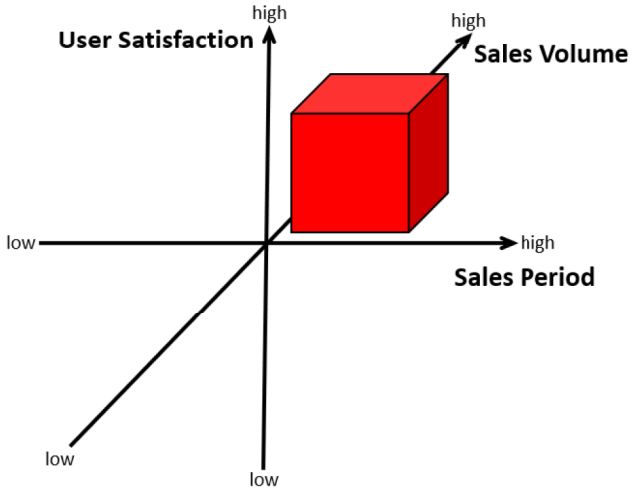



**Figure 2.** Product recommendation based on sales volume, sales period, and user satisfaction.

#### 3.1.1. Sales Volume Scoring

Due to the unavailability of direct sales data on the Rakuten website, we used the number of user reviews as a proxy for sales volume. Previous studies have demonstrated a strong correlation between the number of reviews and the number of purchases [25,26]. Based on this relationship, we calculated a sales volume score ($Vol(a)$) for a target product $a$ using the total number of reviews.

However, we acknowledge the limitations of using review count as a practical proxy for sales volume. Depending on factors such as product type, price, longevity on the platform, and customer engagement patterns, users may be more or less likely to leave reviews. It can introduce bias into sales estimates. For example, products that elicit strong positive or negative opinions, or that are part of incentivized review campaigns, may receive a disproportionate number of reviews. Although we acknowledge the possibility of bias, using review count is commonly accepted and necessary when direct sales data is unavailable. Future work could evaluate methods that incorporate review conversion rates or supplement this proxy with additional behavioral signals to create a more accurate representation of sales velocity.

To classify products, we used the average number of reviews across all products as a threshold value. We chose this data-driven approach because it provides a simple, replicable, and objective baseline for distinguishing between popular and less popular items in our dataset. It is especially crucial since there are no established industry benchmarks for this product category. Products that met or exceeded this threshold value were labeled as having "high sales volume," while those that fell below were labeled as "low sales volume." Using this criterion, all products are appropriately classified based on their sales volume. While a formal sensitivity analysis to test different threshold values was not conducted as part of this study, we acknowledge that exploring the impact of this threshold on model performance is a valuable direction for future work.

#### 3.1.2. Sales Period Scoring

Similar to sales volume, the Rakuten website did not provide direct data on product availability periods. Therefore, we estimated the sales period of each product by calculating the time interval between its first and last review. We used this interval as the estimated sales duration of each product.

We then assigned a sales period score ($Per(a)$) to the target product $a$. Products were classified as having either a "long" or "short" sales period, using the average sales period across all products as the benchmark.

3.1.3. User Satisfaction Scoring

We assessed user satisfaction by combining review sentiment and user ratings. We performed a sentiment analysis on the collected reviews using the Google Cloud Natural Language API (https://cloud.google.com/natural-language) (accessed on 22 June 2023), which assigns sentiment scores ranging from −1.0 (most negative) to 1.0 (most positive). To ensure reproducibility, Table 1 lists the key configuration settings used for the sentiment analysis. Unless otherwise specified, we used the default settings provided by the Google Cloud Natural Language API. The study was conducted in English in document mode using the analyzeSentiment method.

**Table 1.** Google Cloud Natural Language API configuration for sentiment analysis.

| **Parameter** | **Value / Setting** |
|---|---|
| API Method | analyzeSentiment |
| Document Type | PLAIN_TEXT |
| Mode | Sentence-level |
| Language | ja (Japanese) |
| Encoding Type | UTF-8 |
| Sentiment Score Range | −1.0 (most negative) to +1.0 (most positive) |
| Sentiment Magnitude | Included but not used in scoring |
| Threshold for Positive Sentiment | >0.25 |
| Threshold for Neutral Sentiment | $-0.25 \leq$ score $\leq 0.25$ |
| Threshold for Negative Sentiment | <−0.25 |

Based on standard practices in sentiment analysis and the API's documentation, we categorized reviews into three groups: negative (−1.0 to −0.25), neutral (−0.25 to 0.25), and positive (0.25 to 1.0). These thresholds are widely accepted for partitioning sentiment scores into distinct emotional classes. Each review was classified accordingly, and the sentiment score was combined with the user rating to derive a satisfaction score.

We tested four methods of calculating a satisfaction score for each product by combining sentiment scores and user ratings, as determined by a preliminary experiment [27]. The most effective method is to multiply the sentiment score by the user rating for each review and then aggregate these values. The final satisfaction score for the target product $a$ is calculated using sentiment scores and user ratings, as defined in Equation (1):

$$Sat(a) = \frac{\sum_{i=1}^{n} P_i R_i}{\sum_{i=1}^{n} P_i R_i + \sum_{i=1}^{n} |N_i| R_i + \sum_{i=1}^{n} |Z_i| R_i} \quad (1)$$

Here, $P_i$, $N_i$, and $Z_i$ represent the positive, negative, and neutral sentiment scores, respectively, for the $i$th review (with absolute values for $N_i$ and $Z_i$). $R_i$ is the corresponding user rating. The sum is computed over all reviews $n$ for product $a$. Similar to our sales volume scoring, we used the average satisfaction score across all products as a data-driven threshold to categorize products. Products with satisfaction scores above the average were classified as "high satisfaction" and others as "low satisfaction." Using this criterion, all target products were categorized accordingly. Future work could include a sensitivity analysis to evaluate how different satisfaction thresholds might influence the final recommendation rankings.

3.1.4. Product Recommendation Ranking

Traditional recommendation models often rely solely on individual attributes, such as product category, brand, or user ratings. While these methods are beneficial, they may overlook key behavioral and temporal signals. To improve relevance and fairness, we integrated sales volume, sales period, and satisfaction into a unified ranking model.

Since these attributes have different scales, where satisfaction ranges from 0 to 1, sales volume spans a wide numerical range, and sales periods extend from 0 to 365 days, we then normalized each dimension to ensure equal influence. The final recommendation score ($Score(a)$) for the target product $a$ is calculated as shown in Equation (2):

$$Score(a) = Vol(a) \times Per(a) \times Sat(a) \tag{2}$$

Our choice regarding a multiplicative aggregation for the composite score is deliberate. We hypothesize that the top recommended product for a new user should excel in all three dimensions: popularity (sales volume), longevity (sales period), and quality (user satisfaction). A multiplicative approach ensures that products with low scores in any of these areas will be appropriately penalized in the final ranking. For example, a product with high sales volume but poor user satisfaction should not receive a high recommendation. Multiplication naturally captures this 'AND' relationship between attributes. In contrast, an additive model allows a high score in one dimension to compensate for a low score in another.

This composite score reflects a product's market performance, customer satisfaction, and longevity. Products are ranked based on this score to provide new users with balanced recommendations that consider both objective metrics and subjective feedback.

While this multiplicative model provides a robust and interpretable baseline, we acknowledge that its heuristic nature is a limitation. We could explore alternative aggregation methods. For example, a weighted linear combination, $Score(a) = w_1 \cdot Vol(a) + w_3 \cdot Per(a) + w_3 \cdot Sat(a)$, would enable more precise control over the importance of each attribute. Furthermore, incorporating advanced learnable or adaptive weighting mechanisms, such as those based on linear regression or neural networks, could improve performance because the optimal weighting scheme may depend on the user or the context. Exploring such adaptive approaches is a promising direction for future research.

### *3.2. Product Recommendation Method for Existing Users with Purchase Histories*

Collaborative filtering is a widely used strategy for users with an established purchase history. It uses historical data on user–product interactions to improve recommendation accuracy. To further enhance recommendation performance, we propose a bias neural collaborative filtering (BNCF) model. The BNCF model is based on the traditional neural network-based collaborative filtering (NCF) framework and incorporates three product-related bias factors: sales volume, sales period, and user satisfaction. The architecture of the BNCF model is shown in Figure 3.

First, we calculate the bias scores for each product using the methodology described in Section 3.1. Then, we incorporate these scores into the BNCF model alongside the neural feature extraction layers. The BNCF model consists of embedding, hidden, and output layers. These layers are designed to capture latent user–product interactions. Finally, we calculate the recommendation score by combining interaction features and bias-based scores. We train the model using a Bayesian personalized ranking (BPR) loss function, which models user preferences through pairwise ranking and enables personalized optimization. The remainder of this subsection provides a detailed explanation of the BNCF model design and training process.

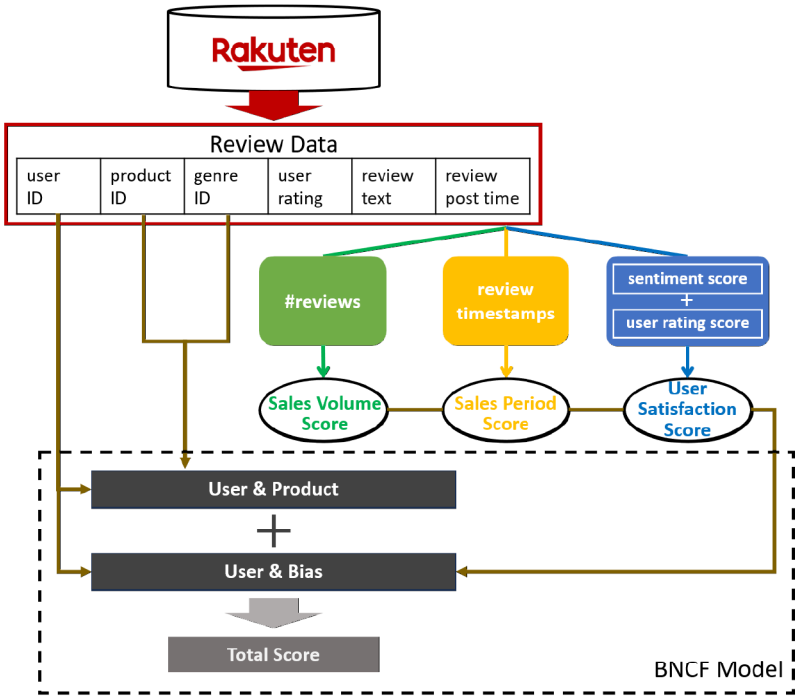



**Figure 3.** Overview of the proposed BNCF model for existing users with purchase histories.

### 3.2.1. Construction of BNCF Model

The neural collaborative filtering (NCF) model employs embedding layers to map users and products onto a shared low-dimensional latent space. It also uses a multilayer perceptron (MLP) to model complex nonlinear interactions. Our BNCF model builds on this by incorporating product and user bias information into the NCF framework, thereby improving the accuracy of personalized recommendations. The architecture is depicted in Figure 4.

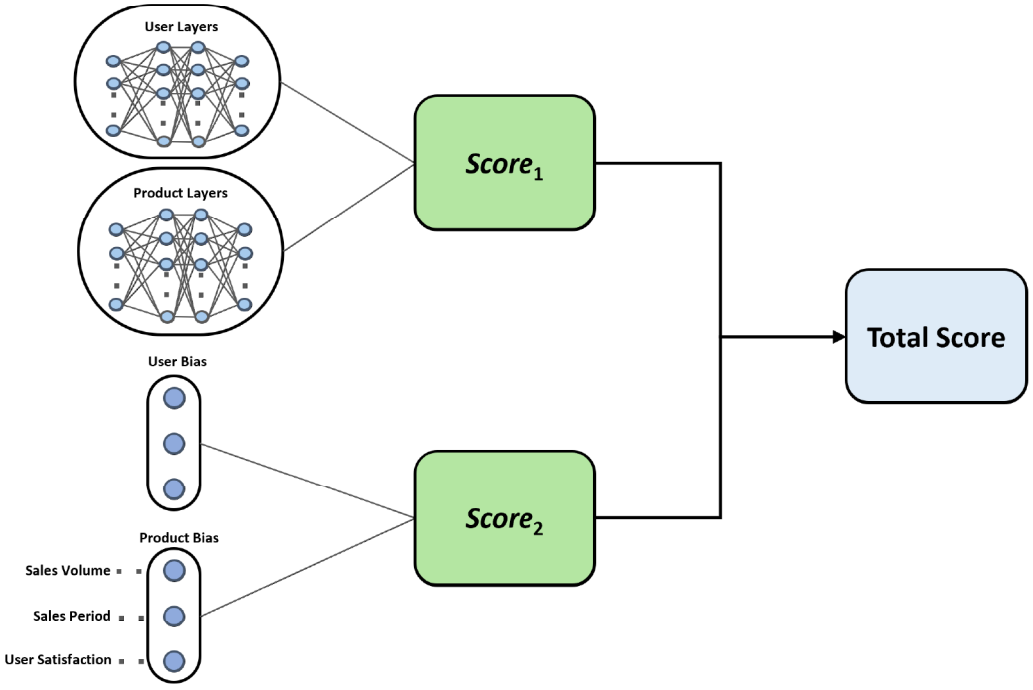



**Figure 4.** BNCF model architecture. The model computes two scores in parallel. $Score_1$ is the collaborative filtering prediction from the NCF module, which processes learned user and item embeddings (represented by the interconnected nodes). $Score_2$ is the bias-aware prediction, derived from the explicit bias feature embeddings (represented by the blue circles).

The BNCF model consists of two main components:

1. **Extraction of User and Product Latent Features**
First, we initialize the user and product embedding vectors. Each vector is $n$-dimensional. Then, we pass these embeddings through two fully connected layers to increase feature expressiveness. Specifically,

$$e_u = \{e_{u_1}, e_{u_2}, e_{u_3}, \ldots, e_{u_n}\}, e_i = \{e_{i_1}, e_{i_2}, e_{i_3}, \ldots, e_{i_n}\}$$

These vectors are processed through two-layer neural networks with ReLU activation to produce refined feature vectors, $h_u$ and $h_i$, which are defined as shown in Equations (3) and (4), respectively:

$$h_u^{(2)} = \text{ReLU}\left(W_u^{(2)} \cdot \text{ReLU}\left(W_u^{(1)} e_u + b_u^{(1)}\right) + b_u^{(2)}\right) \quad (3)$$

$$h_i^{(2)} = \text{ReLU}\left(W_i^{(2)} \cdot \text{ReLU}\left(W_i^{(1)} e_i + b_i^{(1)}\right) + b_i^{(2)}\right) \quad (4)$$

- In our study, $h_u^{(2)}$ and $h_i^{(2)}$ represent the final hidden layer outputs for user $u$ and item $i$, respectively.
- ReLU (rectified linear unit) is a commonly used activation function, defined as $\text{ReLU}(x) = \max(0, x)$.
- $e_u$ and $e_i$ are the embedding vectors for user $u$ and item $i$, respectively.
- $W_u^{(1)}$ and $W_u^{(2)}$ denote the weight matrices for the first and second layers of the user network.
- $b_u^{(1)}$ and $b_u^{(2)}$ represent the bias terms for the first and second layers of the user network.
- $W_i^{(1)}$ and $W_i^{(2)}$ denote the weight matrices for the first and second layers of the item network.
- $b_i^{(1)}$ and $b_i^{(2)}$ represent the bias terms for the first and second layers of the item network.

After obtaining the feature vectors $h_u$ and $h_i$ for the user and the item, respectively, the first interaction score ($Score_1$) is obtained by computing the dot product of these two vectors as shown in Equation (5).

$$Score_1 = h_u^{(2)} \cdot h_i^{(2)} = \sum_{k=1}^{n} h_{u_k} h_{i_k} \quad (5)$$

2. **Incorporation of User and Product Bias Features**
Each user is assigned a three-dimensional bias embedding vector:

$$a_u = \{a_{u_1}, a_{u_2}, a_{u_3}\}$$

Each product is similarly represented using three normalized bias values corresponding to sales volume, sales period, and user satisfaction:

$$a_i = \{a_{i_1}, a_{i_2}, a_{i_3}\}$$

Next, the second interaction score ($Score_2$) is computed by taking the dot product of the user bias vector ($a_u$) and the product bias vector ($a_i$). It reflects the bias in the user–product relationship. It improves the accuracy and flexibility of the model as an additional component of the final recommendation score in Equation (6).

$$Score_2 = a_u \cdot a_i = \sum_{k=1}^{3} a_{u_k} a_{i_k} \tag{6}$$

The final prediction score is the sum of the two components mentioned above, as shown in Equation (7).

$$Total\ Score = Score_1 + Score_2 = \sum_{k=1}^{n} h_{u_k} h_{i_k} + \sum_{k=1}^{3} a_{u_k} a_{i_k} \tag{7}$$

This formulation allows the model to generate more accurate context-aware recommendations by using learned latent features and explicit bias signals.

#### 3.2.2. Bayesian Personalized Ranking Loss

To optimize the BNCF model, we used the Bayesian personalized ranking (BPR) loss function [28], which is suitable for tasks involving implicit feedback recommendations. BPR formulates preference learning as a ranking problem, ensuring that observed (positive) items are ranked higher than unobserved (negative) items for a given user.

The objective of the training is to maximize the difference in preference between positive and negative item pairs using a sigmoid-based loss function. For a given user $u$, a positive item $i$, and a negative item $j$, the BPR loss is defined as in Equation (8):

$$\mathcal{L}_{\text{BPR}} = -\sum_{(u,i,j)\in D} \log \sigma(\hat{x}_{u,i} - \hat{x}_{u,j}) \tag{8}$$

This formulation enables the model to optimize personalized rankings by updating latent factors that directly reflect user preferences.

#### 3.2.3. Model Training

We set the embedding dimensionality to $n = 32$ and incorporated normalized sales volume, sales period, and user satisfaction as bias inputs. During training, we generated user–item triplets (user $u$, positive item $i$, and negative item $j$) and trained the model to satisfy the inequality $\hat{x}_{u,i} > \hat{x}_{u,j}$.

The model was initialized with random parameters and optimized using the Adam optimizer [29]. The hyperparameters were configured as follows:

- Learning rate: 0.001;
- Weight decay: 0.00001;
- Mini-batch size: 256;
- Epochs: 200.

The model parameters were updated using mini-batch stochastic gradient descent and backpropagation until convergence was reached. Figure 5 illustrates the training flow.

To evaluate the model's performance, we conducted experiments under two settings: one that incorporated user bias and one that did not. For each positive sample in the training data (i.e., a purchased product), we generated a negative sample (i.e., a randomly unpurchased product). Training continued until the loss stabilized.

We conducted the evaluation using two widely adopted metrics:

- **NDCG@k (Normalized Discounted Cumulative Gain)** evaluates both the relevance and ranking position of recommended items. We report NDCG@5 and NDCG@10 to measure the quality of the top five and top ten recommendations, respectively.
- **HR@k (Hit Rate)** is the proportion of recommended items with which the user engaged among the top k items. We used HR@5 and HR@10 to evaluate the model's ability to recall recommendations for lists of different sizes.

Combining NDCG and HR provides a more comprehensive evaluation of the recommender system. NDCG focuses on ranking quality, while HR evaluates coverage and alignment with user interests. Together, these metrics offer valuable insights into the system's accuracy, relevance, and user experience.

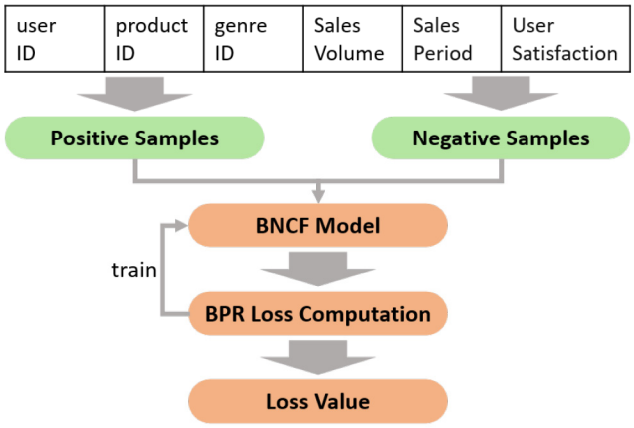



**Figure 5.** BNCF model training flowchart.

## 4. Model Architecture Analysis and Evaluation

In this section, we conducted a systematic evaluation of the proposed model's architecture. The evaluation proceeded in two phases. First, we adjusted the model's structure and parameters and then trained it using a designated training dataset. In the second phase, we evaluated the trained models using a test dataset and conducted a comparative analysis to identify the most effective architecture. First, we implemented a baseline model without item bias and compared it to an extended model that incorporated item bias factors. Based on the results, we experimented with varying the number of negative samples, the number of hidden layers, and the dimensionality of the embedding vectors. We applied these steps to the training and testing procedures to select the optimal model configuration.

### *4.1. Dataset and Preprocessing*

The dataset used in this study was sourced from the Rakuten Ichiba data available in the "Rakuten Dataset," provided by Rakuten Group, Inc. through the IDR Dataset Service of the National Institute of Informatics (NII) [30]. The dataset spans the entirety of 2019 and contains monthly records across five product categories: bags, home appliances, cosmetics, women's fashion, and shoes.

For each product, we extracted relevant information, including product name, product ID, genre ID, user rating, review content, and review timestamp. The dataset also includes user ID and product-level metadata, such as the number of reviews. We applied data cleaning procedures to remove duplicates, null values, and improperly formatted entries.

We constructed two distinct subsets of the dataset to evaluate the two proposed recommendation approaches:

- **Cold-start recommendation (for users without purchase histories):** We selected a sample of 300 products and approximately 18,000 reviews. This subset was used to calculate product-level recommendation scores based on sentiment, sales volume, and sales period, as described in Section 3.1.
- **Personalized recommendation (for users with prior interactions):** We filtered the dataset to include only users who had interacted with at least two products and products purchased by at least two users. The resulting interaction matrix included 4570 users, 4774 products, and 16,241 user–item interactions. To evaluate the BNCF model, we split the dataset by user into training and test sets. One interaction

was randomly selected per user for the test set (706 users), while the remaining were used for training (3723 users). The result is 11,057 training interactions and 706 test interactions.

### *4.2. Experimental Settings*

#### 4.2.1. Experimental Environment

All experiments were conducted on a consistent hardware and software platform to ensure the reproducibility of our results. The specific settings for our experimental environment are detailed in Table 2.

**Table 2.** Experimental hardware and software configuration.

| Category | Component | Specification |
|---|---|---|
| Hardware | CPU | Intel Core i7-12700H @ 2.30 GHz (14 cores) |
| | GPU | NVIDIA GeForce RTX 3060 (6GB VRAM) |
| | Memory (RAM) | 16 GB DDR4 |
| Software | Operating System | Windows 11 Home |
| | Programming Language | Python 3.12.9 |
| | Deep Learning Framework | PyTorch 2.5.0+cu124 |
| | Data Manipulation | Pandas 1.4.2 |
| | Scientific Computing | NumPy 1.21.5 |
| | NLP/Sentiment | Google Cloud Natural Language API |

#### 4.2.2. Hyperparameter Tuning

The BNCF model was initialized with random parameters and optimized using the Adam optimizer [29]. To determine the optimal configuration, we conducted hyperparameter tuning through a grid search strategy. The validation set was created by holding out the last interaction of each user from the training data. Training and validation loss during the model training for 200 epochs.

Table 3 presents the search ranges for each hyperparameter along with the optimal values selected. The final configuration includes an embedding size of 32, 4 negative samples per positive, 2 MLP layers, a learning rate of 0.001, a mini-batch size of 256, and a weight decay of 0.00001.

**Table 3.** BNCF hyperparameter tuning ranges and optimal values.

| Hyperparameter | Search Range | Optimal Value |
|---|---|---|
| Learning Rate | {0.0001, 0.001, 0.01} | 0.001 |
| Embedding Size | {32, 64, 128} | 32 |
| Number of Negative Samples per Positive | {1, 4, 8, 16} | 4 |
| Number of Layers | {1, 2, 3} | 2 |
| Mini-batch Size | {256} | 256 |
| Weight Decay (L2) | {0, 0.00001, 0.0001, 0.001} | 0.00001 |
| Epochs | N/A | 200 |

#### 4.2.3. Loss Function and Training Sample Generation

For model training, we adopted the BPR loss, a pairwise objective function well-suited for implicit feedback datasets. We employed the standard BPR triplet sampling strategy for training both the NCF and BNCF models.

For each observed user–item interaction $(u, i)$, which is treated as a positive sample, we randomly selected one or more items $j$ from the set of items that user $u$ has not interacted with. Each training instance thus forms a triplet $(u, i, j)$, and the model is optimized to maximize the probability that user $u$'s predicted preference score for positive item $i$ exceeds

that for negative item $j$. This negative sampling was performed uniformly and without replacement from the user's set of non-interacted items. The sampling process was repeated dynamically in each training epoch to provide the model with diverse negative examples over time. The number of negative samples generated per positive instance was controlled as a tunable hyperparameter, with the optimal value detailed in our hyperparameter settings (see Table 3).

### *4.3. Accuracy Verification of Product Recommendation Rankings for New Users Without Purchase Histories*

1. **Data Processing**
   We used the same dataset described in Section 4.1, which includes review data for the entire year of 2019. We selected five main product categories, each comprising three subcategories. Within each subcategory, we classified products based on the number of reviews and selected the top 20 products from each subcategory, resulting in a total of 300 products. Next, we performed sentiment analysis on the review texts and calculated three key scores for each product: sales volume score, sales period score, and user satisfaction score. These scores were then binarized ("high" or "low") using threshold-based analysis and normalized. An overall score was subsequently computed to rank products within each subcategory, facilitating the classification and selection of products for further evaluation.
2. **Experimental Method**
   We randomly selected one subcategory from each of the five product categories. Then, we randomly selected 20 products from each subcategory, for a total of 100 products. To evaluate the accuracy of the top ten recommended product rankings, we surveyed twelve university students using a questionnaire. Each participant received comprehensive product information, including the product name, ID, genre ID, user rating, review title and text, sales volume, and sales period. After reviewing this information and the associated reviews, the participants selected their ten preferred products from each subcategory. This questionnaire did not require ethics committee approval as it did not involve animal or human clinical trials and did not raise significant ethical concerns. This study adhered to the ethical principles outlined in the Declaration of Helsinki. All participants were informed of the questionnaire's purpose and provided their informed consent before participating. All participants were guaranteed anonymity and confidentiality, and participation was entirely voluntary. Each product received at least six evaluations. To measure ranking accuracy, we employed three widely used ranking metrics: mean reciprocal rank (MRR), mean average precision (MAP), and normalized discounted cumulative gain at rank 10 (NDCG@10). We computed these metrics using the participants' rankings as the ground truth.
3. **Experimental Results**
   As shown in Table 4, the average MRR across all subcategories is 0.323. The highest MRR is observed in the "Dryer" category (0.450), while the lowest is in "Sweater" (0.142), suggesting suboptimal accuracy for specific product types. The average MAP score is 0.618, with "Suitcase" achieving the highest (0.676) and "Essential Oil" the lowest (0.572), indicating moderate precision in ranking. The average NDCG@10 score is 0.567. The highest score is in the "Suitcase" category (0.623), and the lowest scores are in the "Essential Oil" and "Leather Shoes" categories (0.516 and 0.517, respectively), showing relatively acceptable top-10 ranking performance.

**Table 4.** Ranking accuracy results by rank-aware metrics: MRR, MAP, and NDCG@10.

| Category | MRR | MAP | NDCG@10 |
|---|---|---|---|
| Suitcase | 0.348 | 0.676 | 0.623 |
| Sweater | 0.142 | 0.614 | 0.586 |
| Dryer | 0.450 | 0.637 | 0.593 |
| Essential Oil | 0.263 | 0.572 | 0.516 |
| Leather Shoes | 0.431 | 0.590 | 0.517 |
| AVE | 0.323 | 0.618 | 0.567 |

### *4.4. Comparison of NCF Model with and Without Bias*

To evaluate the effectiveness of incorporating item bias into the proposed BNCF model, we compared it to a baseline NCF model that lacked bias factors. All experiments used 32-dimensional embedding vectors, a 64-dimensional hidden layer, and a 32-dimensional output layer. To ensure a fair comparison, we held the BPR loss function and other hyperparameters constant across models. Each model was trained for 200 epochs. We performed evaluations using NDCG and HR at k = 5 and k = 10. Table 5 summarizes the results.

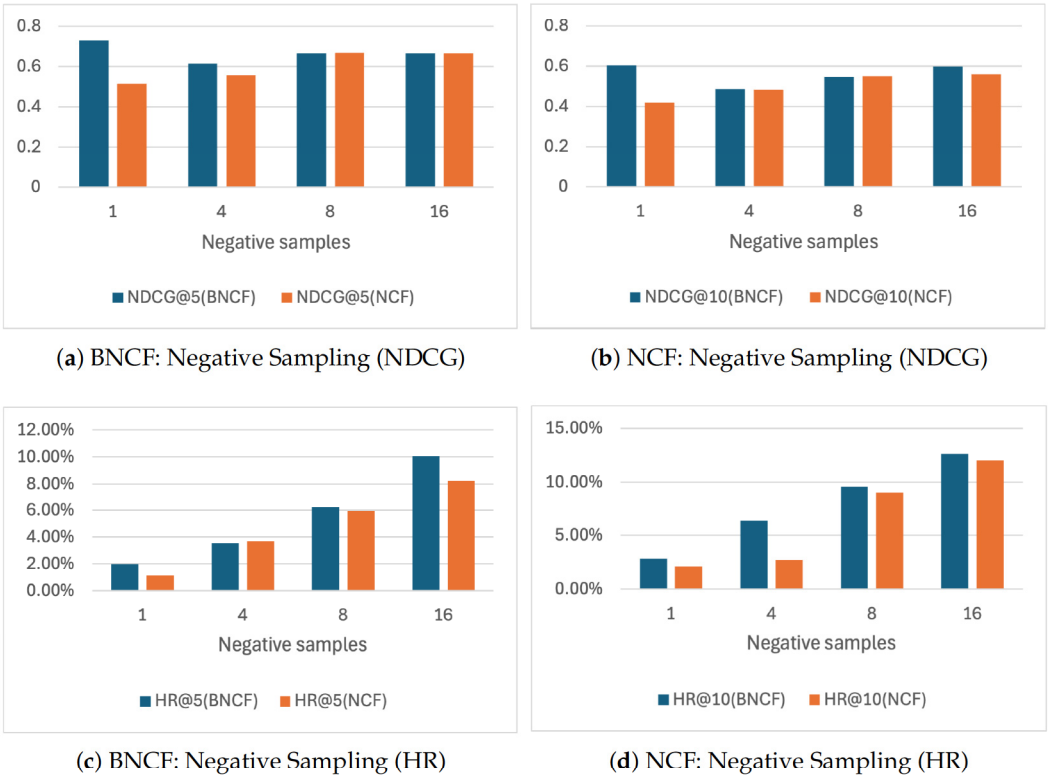
**Table 5.** Performance comparison of NCF with and without bias across varying the number of negative samples.

| Bias | Negative Samples | NDCG@5 | NDCG@10 | HR@5 | HR@10 |
|---|---|---|---|---|---|
| With | 1 | 0.7293 | 0.6023 | 1.98% | 2.83% |
| | 4 | 0.6143 | 0.4856 | 3.54% | 6.37% |
| | 8 | 0.6652 | 0.5444 | 6.23% | 9.63% |
| | 16 | 0.6645 | 0.5981 | 10.06% | 12.61% |
| Without | 1 | 0.5136 | 0.4200 | 1.13% | 2.12% |
| | 4 | 0.5576 | 0.4841 | 3.68% | 2.69% |
| | 8 | 0.6685 | 0.5500 | 5.95% | 9.07% |
| | 16 | 0.6648 | 0.5590 | 8.22% | 12.04% |

As shown in Figure 6, increasing the number of negative samples improves both NDCG and HR metrics for both models. This improvement stems from the model's enhanced ability to learn from more negative examples, which strengthens its generalization capability. However, using too many negative samples can significantly increase training time and the risk of overfitting. It highlights the importance of balancing performance with computational cost, which is crucial for achieving optimal efficiency and reducing resource consumption.

The BNCF model outperforms the NCF model when negative samples are scarce. It is evident from its higher NDCG@5 and NDCG@10 scores. It suggests that incorporating bias factors provides contextual information, enabling the model to capture latent user–item interactions accurately. As the number of negative samples increases, however, the performance gap between the two models narrows, although the BNCF model still performs slightly better.

On the other hand, the unbiased NCF model shows more consistent improvement with negative samples. It suggests that it relies more on interaction data. Larger sample sizes allow the model to more accurately identify user preferences, thereby improving hit rates and ranking quality. However, in this scenario, the relative contribution of item bias decreases while model complexity increases, which could lead to overfitting.




(**a**) BNCF: Negative Sampling (NDCG)

(**b**) NCF: Negative Sampling (NDCG)

(**c**) BNCF: Negative Sampling (HR)

(**d**) NCF: Negative Sampling (HR)

**Figure 6.** Comparison of NDCG and HR metrics for BNCF and NCF models across different numbers of negative samples (1, 4, 8, and 16).

Therefore, we recommend using the BNCF model when the number of negative samples is limited as the bias factors significantly improve performance. For large-scale negative sampling with sufficient computational resources, the simpler NCF model without bias can maintain competitive performance while reducing the risk of overfitting and training overhead.

### *4.5. Comparison with Baseline Models*

To validate the effectiveness of our proposed BNCF model, we conducted a comparative evaluation against three baseline models: the standard NCF, a classic model with explicit bias terms (BiasedMF) [16], and a representative state-of-the-art model based on graph neural networks (LightGCN) [31].

- **NCF:** This model serves as our foundational baseline. Since it does not incorporate any form of bias modeling, it is used to isolate the performance gains achieved by our proposed bias integration.
- **BiasedMF:** It is a classical matrix factorization model that incorporates scalar user and item bias terms. This model provides a basis for evaluating the effectiveness of explicit bias modeling in traditional recommender systems.
- **LightGCN:** This model represents a state-of-the-art approach based on graph convolutional networks. It captures high-order connectivity within the user–item interaction graph and is widely recognized for its strong performance in recommendation tasks.

All baseline models were implemented using publicly available libraries. Their hyperparameters were optimized using our validation set to ensure a fair and meaningful comparison.

As shown in Table 6, the standard NCF model achieves an NDCG@10 of 0.6209 and an HR@10 of 8.07%. Its lack of bias modeling limits its ability to personalize recommendations

based on item-specific factors. While BiasedMF introduces explicit bias factors, it relies on linear interactions, and it cannot capture complex user–item dynamics. Consequently, its performance is substantially lower. The NDCG@10 is 0.4779. The HR@10 is only 1.98%.

**Table 6.** Performance comparison of BNCF and baseline models.

| Model | NDCG@10 | HR@10 | Description |
|---|---|---|---|
| NCF | 0.6209 | 8.07% | Base model without any bias component |
| BiasedMF | 0.4779 | 1.98% | Classic model with explicit bias terms |
| LightGCN | 0.9721 | 19.73% | State-of-the-art GNN-based model |
| BNCF (ours) | 0.9848 | 19.69% | Proposed model with multi-faceted bias |

In contrast, LightGCN uses the user–item graph structure and propagates latent preferences through multiple graph convolution layers. This model demonstrates strong overall performance, achieving an NDCG@10 of 0.9721 and an HR@10 of 19.73%.

Our proposed BNCF model combines the expressiveness of deep neural networks with domain-aware bias integration. By incorporating normalized sales volume, sales period, and user satisfaction scores into the embedding space, the BNCF model effectively captures both business-driven signals and nonlinear user–item interactions. The model achieves the highest NDCG@10 of 0.9848, indicating superior ranking precision. Its HR@10 reaches 19.69%, which is comparable to the performance of LightGCN.

These results confirm the advantage of combining multidimensional bias information with deep collaborative filtering frameworks. The BNCF model provides personalized accurate recommendations, outperforming traditional matrix factorization methods and advanced graph-based models in terms of ranking accuracy.

### *4.6. Ablation Study of Bias Components*

To assess the individual contribution of each proposed bias component in the BNCF model, we conducted an ablation study. Specifically, we created three variants of the model by selectively removing one bias factor at a time. These factors include sales volume (Vol), sales period (Per), and user satisfaction (Sat). Each variant retains the remaining two components. For comparative purposes, we also included the standard NCF model, which excludes all bias factors entirely.

Each model variant was trained under the identical configuration detailed in Section 4.2, with a mini-batch size of 256 and 4 negative samples per positive instance. The performance results, evaluated using NDCG@10 and HR@10, are presented in Table 7.

**Table 7.** Ablation study results of bias components.

| Bias Configuration | NDCG@10 | HR@10 |
|---|---|---|
| Full BNCF (Vol, Per, Sat) | 0.9848 | 19.69% |
| Without Satisfaction (Vol, Per) | 0.9773 | 19.83% |
| Without Sales Period (Vol, Sat) | 0.9937 | 19.41% |
| Without Sales Volume (Per, Sat) | 0.9831 | 19.41% |
| No Bias (Standard NCF) | 0.6209 | 8.07% |

The results in Table 7 provide valuable insights into the role of each bias component. All model variants that incorporate bias factors outperform the standard NCF model. These results confirm the effectiveness of incorporating product-level attributes into the recommendation process.

Among the three components, sales volume (Vol) has the strongest influence on hit rate. Removing this factor causes HR@10 to decline from 19.69% to 19.41%. This observation

indicates that product popularity plays a critical role in enhancing user engagement. In contrast, user satisfaction (Sat) has the most significant effect on ranking precision. Its removal results in a decrease in NDCG@10 from 0.9848 to 0.9773, while HR@10 slightly increases. It suggests a shift toward recommending more popular but potentially less relevant items.

The sales period (Per) component exhibits the least impact on both metrics. Its removal leads to a slight increase in NDCG@10, suggesting that this feature may provide limited additional value when sales volume (Vol) and user satisfaction (Sat) are already present.

In summary, this ablation study shows that the full BNCF model provides the most balanced and effective performance. Sales volume (Vol) contributes to exposure and engagement, while user satisfaction (Sat) enhances personalized ranking. The combination of these diverse factors enables the model to provide recommendations that are both accurate and user-centered.

### *4.7. Analysis of Hidden Layer Structures in BNCF Model*

To enhance the model's capacity for capturing intricate user–item interactions, we investigated the impact of various hidden layer architectures on the performance of the BNCF model. Specifically, we modified the original NCF model by introducing additional hidden layers to evaluate their effect. We defined the baseline architecture (Model A) as 32–64–32. We created three enhanced variants: Model B had one additional hidden layer, and Model C had two additional hidden layers. The bias component structure remained unchanged across all models.

These modifications were made based on the idea that deeper architectures can better capture complex nonlinear interactions, which could lead to richer feature representations and improved generalization. To analyze the interaction between architecture depth and negative sampling, we evaluated each model using 1, 4, and 8 negative samples.

As shown in Table 8, increasing the model depth results in clear performance gains. For example, Model B achieved significant improvements over Model A when using one negative sample: NDCG@5 and NDCG@10 increased by 24.31% and 46.14%, respectively. In terms of hit rate, HR@5 and HR@10 increased approximately eightfold and 5.56-fold, respectively. These results suggest that the added hidden layer in Model B significantly enhances the model's expressiveness.

**Table 8.** Performance metrics of different models.

| Model | Negative Samples | NDCG@5 | NDCG@10 | HR@5 | HR@10 |
|---|---|---|---|---|---|
| A | 1 | 0.7293 | 0.6023 | 1.98% | 2.83% |
| | 4 | 0.6143 | 0.4856 | 3.54% | 6.37% |
| | 8 | 0.6652 | 0.5444 | 6.23% | 9.63% |
| B | 1 | 0.9066 | 0.8802 | 17.71% | 18.56% |
| | 4 | 0.9946 | 0.9848 | 19.41% | 19.69% |
| | 8 | 0.9884 | 0.9742 | 19.55% | 19.67% |
| C | 1 | 0.8744 | 0.8334 | 16.15% | 17.42% |
| | 4 | 0.9560 | 0.9234 | 18.27% | 19.26% |
| | 8 | 0.8886 | 0.8675 | 18.27% | 18.98% |

As shown in Figure 7, increasing the number of negative samples from 1 to 4 substantially improves performance. However, the marginal benefit diminishes when the number of negative samples increases from 4 to 8. Furthermore, models with excessive depth may encounter training instability. For example, Model C exhibited numerical instability during loss computation, particularly at higher negative sampling rates. This predicament

is ubiquitous in deep networks as vanishing or exploding gradients can compromise the integrity of the training process. As suggested by Kloberdanz et al. [32], introducing a small epsilon (e.g., $1 \times 10^{-9}$) into the logarithmic function can mitigate such numerical issues, particularly avoiding undefined operations like $\log(0)$.




(**a**) Model A (**b**) Model B




(**c**) Model C (**d**) Model D

**Figure 7.** Performance of different models under various negative sampling conditions.

Among the three variants, Model B consistently delivers the best overall performance, especially when using four or more negative samples. Model C, while deeper, does not provide significant performance gains over Model B and introduces greater computational and numerical complexity. Model A, as the baseline, shows the weakest performance.

These results indicate that Model B (32–64–64–32) offers an effective trade-off between expressiveness and computational efficiency. Its moderate complexity allows it to model nonlinear user–item interactions effectively without incurring the training difficulties observed in deeper models. Moreover, using four negative samples provides a practical balance between recommendation performance and computational cost. Based on these findings, we conclude that Model B with four negative samples is the optimal configuration for our BNCF model.

### *4.8. Analysis of Embedding Dimensions in BNCF Model*

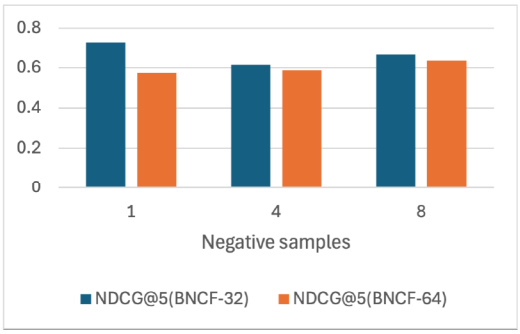
We investigated how the size of the embedding dimension affects the performance of the BNCF model. The original model, BNCF-32, used embeddings with dimensions of 32, 64, and 32. Next, we constructed a larger variant, BNCF-64, with dimensions of 64, 128, and 64. Both models included item bias terms of equal size. We conducted experiments using 1, 4, and 8 negative samples.

As illustrated in Table 9 and Figure 8, increasing the embedding dimension did not lead to performance improvement. BNCF-32 outperformed BNCF-64 across all evaluation metrics. For example, when using one negative sample, BNCF-32 achieved significantly higher NDCG@5 (0.7293 vs. 0.5725) and HR@5 (1.98% vs. 1.27%). Even when using

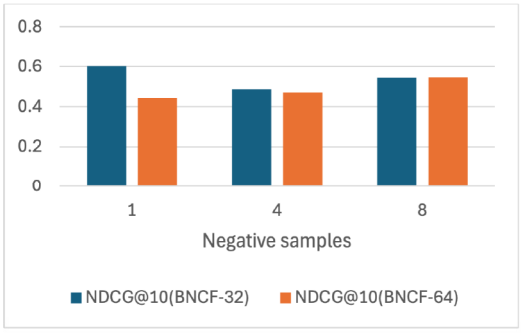
eight negative samples, the improvement in BNCF-64 remained marginal and did not surpass BNCF-32.

**Table 9.** Performance metrics of different BNCF models.

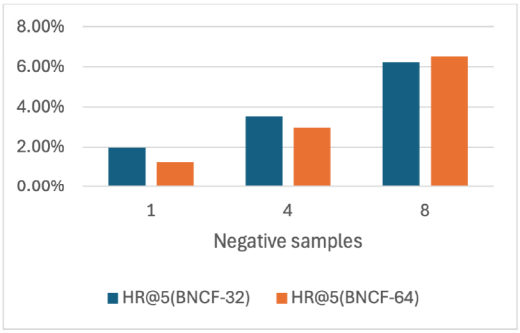
| Model | Negative Samples | NDCG@5 | NDCG@10 | HR@5 | HR@10 |
|---|---|---|---|---|---|
| BNCF-32 | 1 | 0.7293 | 0.6023 | 1.98% | 2.83% |
| | 4 | 0.6143 | 0.4856 | 3.54% | 6.37% |
| | 8 | 0.6652 | 0.5444 | 6.23% | 9.63% |
| BNCF-64 | 1 | 0.5725 | 0.4430 | 1.27% | 2.69% |
| | 4 | 0.5881 | 0.4709 | 2.97% | 5.24% |
| | 8 | 0.6359 | 0.5454 | 6.52% | 9.21% |



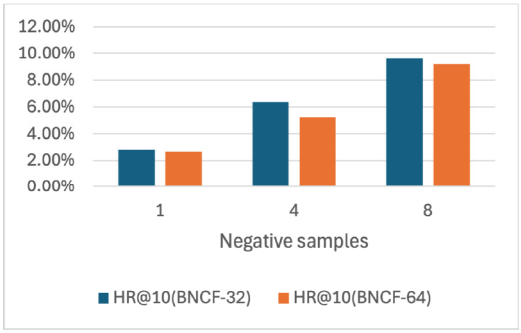

(**a**) NDCG (BNCF-32)




(**b**) NDCG (BNCF-64)




(**c**) HR (BNCF-32)




(**d**) HR (BNCF-64)

**Figure 8.** Performance comparison of BNCF-32 and BNCF-64 across NDCG and HR metrics.

Several factors may explain this degradation. First, increasing embedding dimensions substantially enlarges the parameter space, which raises the risk of overfitting, particularly in scenarios with limited training data. Second, complex models are more sensitive to noise and require more robust regularization. Third, deep models with large embeddings are susceptible to vanishing or exploding gradients, which complicates the training process and can lead to suboptimal convergence.

In contrast, BNCF-32 provides a well-balanced configuration: its complexity is sufficient to capture meaningful user–item interaction patterns without introducing the instability and overfitting risks associated with larger embeddings. The results suggest that further enlarging embedding dimensions beyond a certain threshold does not guarantee improved performance and may hinder generalization.

In conclusion, the original BNCF-32 architecture demonstrates superior robustness and efficiency compared to its higher-dimensional counterpart. It effectively balances representational power, training stability, and computational cost, making it the preferred choice for practical applications.

## 5. Discussion

In Sections 3.1 and 3.2, we introduced two distinct recommendation methods. Despite the differences in dataset composition and experimental design, a comparison of the NDCG metrics allows us to assess and contrast their effectiveness. We refer to the recommendation approach for users with no purchase histories as Method 1 and the approach tailored for users with prior purchase behavior as Method 2.

### *5.1. Method 1: Rule-Based Recommendation for Cold-Start Users*

In Method 1, we focused on five product categories: suitcases, sweaters, dryers, essential oils, and leather shoes, selecting 20 items per category for a total of 100 products. Recommendations were generated by ranking the products based on a composite score derived from sales volume, sales duration, and user satisfaction.

The experimental results showed an average NDCG@10 of 0.567, with the suitcase category achieving the highest score (0.623). Categories such as essential oils (0.516) and leather shoes (0.517) exhibited lower performance. Upon analysis, we attribute this disparity to the uniform weighting of the factors in our scoring function. Although the method assumes that users value sales volume, sales period, and satisfaction equally, user preferences vary significantly. Some users prioritize longevity in sales (sales period), while others focus more on recent popularity (sales volume) or product quality (user satisfaction). This variability leads to a divergence between system-generated rankings and user-perceived relevance.

For example, a product with a high sales volume, a short sales duration, and high customer satisfaction is likely to be popular and widely accepted. These attributes attract users. However, if the model continues to use a static weighting formula, it may misrepresent user intent, resulting in less precise recommendations. Evaluation metrics such as MAP and NDCG@10 reflect this issue. These metrics typically range between 0.5 and 0.6, suggesting that nearly half of recommendations deviate from user preferences.

### *5.2. Method 2: Deep Learning-Based BNCF for Personalized Recommendation*

In Method 2, we employed a larger and more complex dataset consisting of 4570 users, 4774 items, and 16,241 records of user–item interactions. We proposed the BNCF model, which integrates three item-level bias features—sales volume, sales period, and user satisfaction—into the standard NCF architecture. After testing, we randomly selected the bias embedding vector parameters of 20 users from the model that showed the best performance and mapped them to the 0–1 interval using the sigmoid function. Figure 9 shows the relationship between the three bias parameters of the users.

Under different configurations of hidden layers and negative sampling rates (1, 4, and 8), the model demonstrated strong performance. For instance, with one negative sample per user, NDCG@10 reached 0.8802; with four negative samples, it increased further to 0.9848. These results are markedly superior to those obtained in Method 1.

Several factors contributed to this performance improvement:

1. **Rich and Diverse Dataset:**
   Method 2 benefits from a significantly larger dataset, allowing the model to learn more nuanced patterns of user preferences and item interactions, which improves the generalization and accuracy of recommendations.
2. **Advanced Model Architecture:**
   Method 2 uses a neural collaborative filtering framework to capture the complex nonlinear interactions between users and items. It is a significant advantage over Method 1's rule-based approach as it cannot adapt to intricate user behavior.
3. **Integration of Bias Features:**

The inclusion of item-specific bias factors enhanced the model's contextual understanding of products. By embedding attributes such as popularity and user satisfaction directly into the model, the BNCF approach provided a richer and more personalized recommendation output.

Collectively, these factors illustrate the importance of employing sophisticated models and multidimensional item data to improve the quality of recommendations.

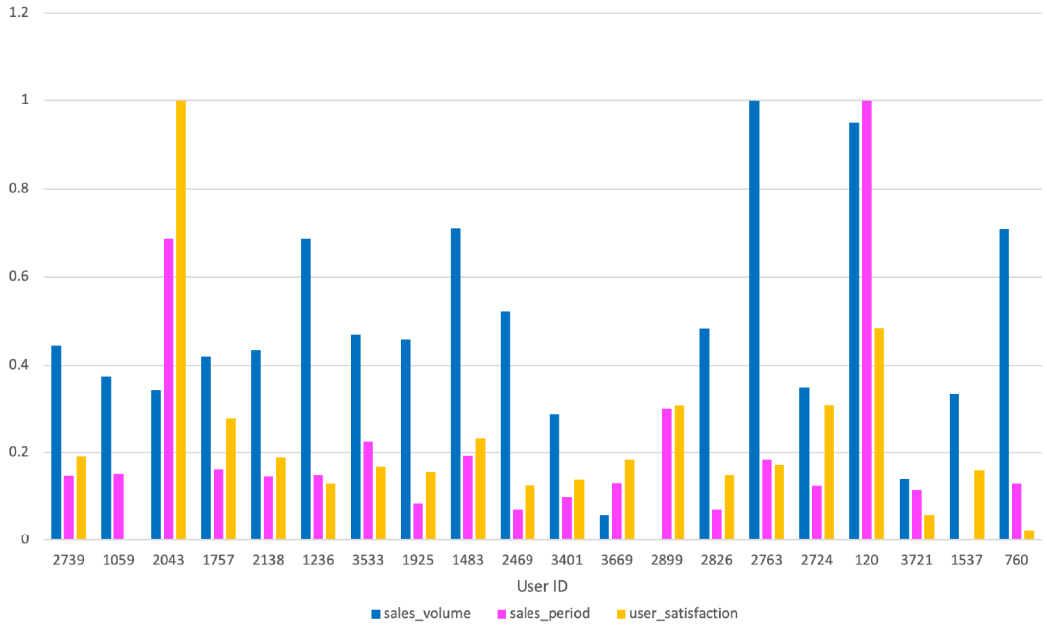



**Figure 9.** Bias parameter visualization of BNCF model.

### *5.3. Comparative Model Analysis Within Method 2*

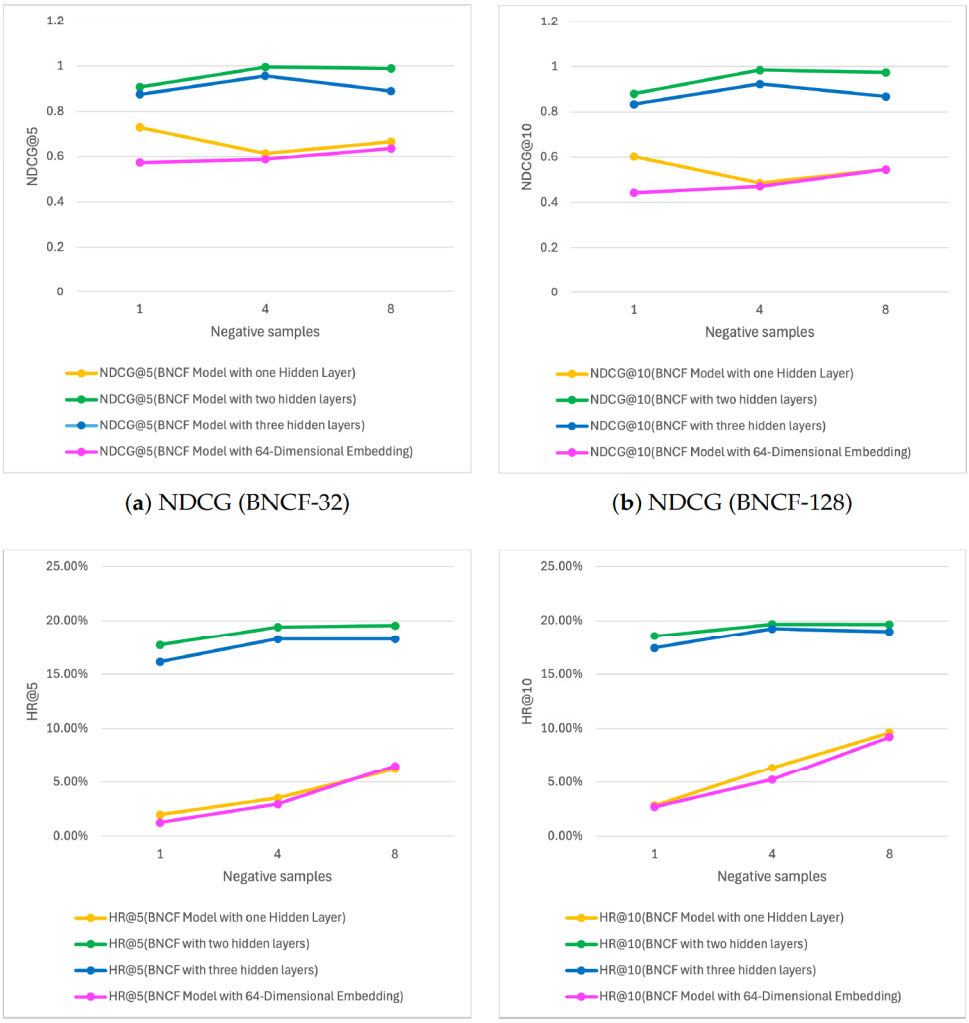
To further refine Method 2, we conducted comparative experiments by varying the architecture of the BNCF model. Specifically, we tested different hidden-layer configurations and embedding dimensions. The goal was to identify an optimal balance between model complexity, expressive capacity, and generalization performance.

Among the configurations tested, the 32–64–64–32 architecture consistently outperformed the others, as indicated in Figure 10. This model achieved the best overall performance across both the NDCG and HR metrics. Compared to the simpler 32–64–32 structure or the larger 64–128–64 configuration, this architecture provided enhanced feature representation without incurring significant computational overhead or overfitting.

Our experiments also underscored the importance of the following:

- **Managing Numerical Stability:** During training, we encountered instability issues in deeper models (e.g., gradient explosion/vanishing). Mitigating these through regularization techniques and loss function smoothing (e.g., adding small constants to avoid $\log(0)$) proved essential for stable optimization.
- **Selecting Optimal Negative Sampling Ratios:** We found that increasing the number of negative samples to four substantially improved performance. However, further increases yielded diminishing returns and increased computational cost. Thus, using four negative samples offered a practical trade-off between performance and efficiency.

In summary, Method 2 demonstrates that thoughtful architectural design, feature integration, and training strategies can dramatically enhance recommendation performance. In real-world applications, the combination of deep learning techniques, multidimensional item features, and user interaction data offers a robust solution to the challenges faced by modern recommender systems.




(**a**) NDCG (BNCF-32)

(**b**) NDCG (BNCF-128)

(**c**) HR (BNCF-32)

(**d**) HR (BNCF-128)

**Figure 10.** Comparison of performance metrics (NDCG and HR) for BNCF models with different structures: BNCF-32 and BNCF-128.

## 6. Conclusions

In this study, we proposed two recommendation strategies tailored to different user scenarios: one for users without purchase histories and another for users with prior interactions. For users without purchase histories, we developed a rule-based recommendation method grounded in three key product attributes: sales volume, sales period, and user satisfaction. We quantified user satisfaction through sentiment analysis combined with rating scores. Then, we derived a composite score through threshold analysis to rank products. We evaluated the system's effectiveness using product review data from Rakuten Ichiba. The results showed that the method could identify and recommend products with high user satisfaction, strong sales, and extended sales cycles. However, it failed to align with individual user preferences due to the lack of personalized behavioral data.

To address this limitation, we proposed a BNCF model for users with purchase histories. The BNCF model extends the standard neural collaborative filtering framework by incorporating item-level bias factors, specifically sales volume, sales period, and user satisfaction, into the recommendation process. We conducted extensive experiments to examine the impact of various conditions, including the inclusion of bias factors, different numbers of negative samples, variations in model depth, and embedding dimensionality. These experiments led to the identification of an optimal BNCF configuration, which

demonstrated superior performance in both ranking quality and accuracy compared to the baseline models.

The comparative analysis of the two methods provided several key insights. Users without purchase histories present a significant challenge due to the absence of interaction data, which restricts the model to general recommendations based solely on product-level attributes. As a result, the capacity to deliver personalized and precise recommendations is considerably limited. In contrast, users with purchase histories provide valuable behavioral data that we can use to model intricate user–item interactions. The BNCF model uses deep learning techniques. These techniques identify latent relationships. The result is more precise and customized recommendations. The experimental results confirmed that this approach significantly improves recommendation accuracy and user satisfaction.

Future research will aim to expand the dataset and improve product classification granularity for users without purchase histories. We plan to represent product features within a three-dimensional space and introduce a dynamic weighting mechanism to account for variations in user preference priorities.

For users with prior interactions, future efforts will focus on enhancing the BNCF framework through the integration of the following techniques:

- Graph neural networks (GNNs) can model complex relationships among users and items across different types of interactions.
- Cross-modal attention mechanisms effectively fuse multiple data sources, including text, images, and numbers.
- Residual connections can stabilize training and enhance the performance of deeper network architectures.

These advancements are expected to improve the accuracy, robustness, and adaptability of recommender systems in real-world applications.

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

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Ethical review and approval were waived for this study due to its nature as an anonymous user questionnaire survey, which does not involve personal data collection or invasive procedures, in accordance with the "Ethical Guidelines for Medical and Biological Research Involving Human Subjects" established by the Japanese Ministry of Health, Labour, and Welfare (MHLW) and the "Regulations for General Research Involving Human Subjects" at Yamaguchi University.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data analyzed in this study is subject to the following licenses and restrictions: National Institute of Informatics's Informatics Research Data Repository for Research Purposes. Requests to access this dataset should be directed to https://rit.rakuten.com/data_release/ (accessed on 21 May 2023). Further inquiries can be directed to the corresponding authors.

**Acknowledgments:** In this paper, the authors used the "Rakuten Dataset" (https://rit.rakuten.com/data_release/) (accessed on 21 May 2023) provided by Rakuten Group, Inc. via IDR Dataset Service of the National Institute of Informatics.

**Conflicts of Interest:** Author Haoyang Xia is employed by MICWARE Co., Ltd. However, this research was conducted independently as part of his doctoral studies at Yamaguchi University and was not supported by MICWARE Co., Ltd. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| BiasedMF | Biased matrix factorization |
| BNCF | Bias neural collaborative filtering |
| BPR | Bayesian personalized ranking |
| CF | Collaborative filtering |
| CNNs | Convolutional neural networks |
| DeepCoNNs | Deep cooperative neural networks |
| DNNs | Deep neural networks |
| GCNs | Graph convolutional networks |
| GNNs | Graph neural networks |
| HR | Hit rate |
| KNN | K-nearest neighbor |
| LSTM | Long short-term memory |
| MAP | Mean average precision |
| MRR | Mean reciprocal rank |
| NDCG | Normalized discounted cumulative gain |
| NCF | Neural collaborative filtering |
| NII | National Institute of Informatics |
| NLP | Natural language processing |
| RC-DFM | Review and content-based deep fusion model |
| ReLU | Rectified linear unit |

## References

1. Lackermair, G.; Kailer, D.; Kanmaz, K. Importance of online product reviews from a consumer's perspective. *Adv. Econ. Bus.* **2013**, *1*, 1–5. [CrossRef]
2. Changchit, C.; Klaus, T. Determinants and impact of online reviews on product satisfaction. *J. Internet Commer.* **2020**, *19*, 82–102. [CrossRef]
3. Im, I.; Hars, A. Does a one-size recommendation system fit all? The effectiveness of collaborative filtering based recommendation systems across different domains and search modes. *ACM Trans. Inf. Syst. (TOIS)* **2007**, *26*, 4-es. [CrossRef]
4. Çano, E.; Morisio, M. Hybrid recommender systems: A systematic literature review. *Intell. Data Anal.* **2017**, *21*, 1487–1524. [CrossRef]
5. Sarwar, B.; Karypis, G.; Konstan, J.; Riedl, J. Item-based Collaborative Filtering Recommendation Algorithms. In Proceedings of the WWW'01: 10th International Conference on World Wide Web, HIong Kong, China, 1–5 May 2001; pp. 285–295. [CrossRef]
6. Ko, H.; Lee, S.; Park, Y.; Choi, A. A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields. *Electronics* **2022**, *11*, 141. [CrossRef]
7. Bobadilla, J.; Dueñas-Lerín, J.; Ortega, F.; Gutierrez, A. Comprehensive evaluation of matrix factorization models for collaborative filtering recommender systems. *Int. J. Interact. Multimed. Artif. Intell.* **2024**, *8*, 15–23. [CrossRef]
8. Adeniyi, D.; Wei, Z.; Yongquan, Y. Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method. *Appl. Comput. Inform.* **2016**, *12*, 90–108. [CrossRef]
9. Zou, C.; Zhang, D.; Wan, J.; Hassan, M.M.; Lloret, J. Using concept lattice for personalized recommendation system design. *IEEE Syst. J.* **2017**, *11*, 305–314. [CrossRef]
10. Wu, J.; Rao, H.; Fan, X.; Wang, Y. A multidimensional application recommender system based on user feature clustering and contextual information. *J. Integr. Technol.* **2021**, *10*, 22. [CrossRef]
11. He, X.; Liao, L.; Zhang, H.; Nie, L.; Hu, X.; Chua, T.S. Neural Collaborative Filtering. In Proceedings of the WWW'17: 26th International Conference on World Wide Web, Perth, Australia, 3–7 April 2017; pp. 173–182. [CrossRef]

12. Messaoudi, F.; Loukili, M. E-commerce personalized recommendations: A deep neural collaborative filtering approach. *SN Oper. Res. Forum* **2024**, *5*, 5. [CrossRef]
13. Cheng, H.T.; Koc, L.; Harmsen, J.; Shaked, T.; Chandra, T.; Aradhye, H.; Anderson, G.; Corrado, G.; Chai, W.; Ispir, M.; et al. Wide & Deep Learning for Recommender Systems. In Proceedings of the DLRS 2016: 1st Workshop on Deep Learning for Recommender Systems, Boston, MA, USA, 15 September 2016; pp. 7–10. [CrossRef]
14. Patel, R.; Thakkar, P.; Ukani, V. CNNRec: Convolutional neural network based recommender systems—A survey. *Eng. Appl. Artif. Intell.* **2024**, *133*, 108062. [CrossRef]
15. Daneshvar, H.; Ravanmehr, R. A social hybrid recommendation system using LSTM and CNN. *Concurr. Comput. Pract. Exp.* **2022**, *34*, e7015. [CrossRef]
16. Koren, Y.; Bell, R.; Volinsky, C. Matrix Factorization Techniques for Recommender Systems. *Computer* **2009**, *42*, 30–37. [CrossRef]
17. Redhu, S.; Srivastava, S.; Bansal, B.; Gupta, G. Sentiment analysis using text mining: A review. *Int. J. Data Sci. Technol.* **2018**, *4*, 49–53. [CrossRef]
18. Pham, T.D.; Vo, D.; Li, F.; Baker, K.; Han, B.; Lindsay, L.; Pashna, M.; Rowley, R. Natural language processing for analysis of student online sentiment in a postgraduate program. *Pac. J. Technol. Enhanc. Learn.* **2020**, *2*, 15–30. [CrossRef]
19. Zheng, L.; Noroozi, V.; Yu, P.S. Joint Deep Modeling of Users and Items Using Reviews for Recommendation. In Proceedings of the WSDM'17: Tenth ACM International Conference on Web Search and Data Mining, Cambridge, UK, 6–10 February 2017; pp. 425–434. [CrossRef]
20. Fu, W.; Peng, Z.; Wang, S.; Xu, Y.; Li, J. Deeply Fusing Reviews and Contents for Cold Start Users in Cross-Domain Recommendation Systems. *Proc. AAAI Conf. Artif. Intell.* **2019**, *33*, 94–101. [CrossRef]
21. Xu, K.; Zhou, H.; Zheng, H.; Zhu, M.; Xin, Q. Intelligent classification and personalized recommendation of E-commerce products based on machine learning. *Appl. Comput. Eng.* **2024**, *64*, 147–153. [CrossRef]
22. Bellar, O.; Baina, A.; Ballafkih, M. Sentiment analysis: Predicting product reviews for E-commerce recommendations using deep learning and transformers. *Mathematics* **2024**, *12*, 2403. [CrossRef]
23. Xiang, A.; Huang, B.; Guo, X.; Yang, H.; Zheng, T. A Neural Matrix Decomposition Recommender System Model based on the Multimodal Large Language Model. *arXiv* **2024**, arXiv:2407.08942. [CrossRef]
24. Poriya, A.; Bhagat, T.; Patel, N.; Sharma, R. Non-Personalized Recommender Systems and User-based Collaborative Recommender Systems. *Int. J. Appl. Inf. Syst.* **2014**, *6*, 22–27. [CrossRef]
25. Xian, Y.; Yanzhong, D.; Jiangning, W. Analysis on the relationship between the number of product reviews and the number of purchases based on symbolic regression. *J. Syst. Eng.* **2020**, *35*, 289–300. (In Chinese) [CrossRef]
26. Lee, J.; Lee, J.N.; Shin, H. The long tail or the short tail: The category-specific impact of eWOM on sales distribution. *Decis. Support Syst.* **2011**, *51*, 466–479. [CrossRef]
27. Xia, H.; Wang, Y. A Product Recommendation Method by Analyzing Sales Volume, Sales Period, and User Satisfaction. In Proceedings of the 2024 12th International Conference on Information and Education Technology (ICIET), Yamaguchi, Japan, 18–20 March 2024; pp. 407–411. [CrossRef]
28. Rendle, S.; Freudenthaler, C.; Gantner, Z.; Schmidt-Thieme, L. BPR: Bayesian Personalized Ranking from Implicit Feedback. In Proceedings of the UAI'09: Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, Montreal, QC, Canada, 18–21 June 2009.
29. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980. [CrossRef]
30. Rakuten Group, Inc. Rakuten Ichiba Data. Informatics Research Data Repository, National Institute of Informatics. 2020. (Dataset). Available online: https://doi.org/10.32130/idr.2.1 (accessed on 21 May 2023).
31. He, X.; Deng, K.; Wang, X.; Li, Y.; Zhang, Y.; Wang, M. LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation. In Proceedings of the SIGIR'20: 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, China, 25–30 July 2020; pp. 639–648. [CrossRef]
32. Kloberdanz, E.; Kloberdanz, K.G.; Le, W. DeepStability: A Study of Unstable Numerical Methods and Their Solutions in Deep Learning. In Proceedings of the ICSE'22: 44th International Conference on Software Engineering, Pittsburgh, PA, USA, 21–29 May 2022; pp. 586–597. [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.

electronics

Article

# Multi-Sensor Fusion for Autonomous Mobile Robot Docking: Integrating LiDAR, YOLO-Based AprilTag Detection, and Depth-Aided Localization

**Yanyan Dai and Kidong Lee ***

Robotics Department, Yeungnam University, Gyeongsan 38541, Republic of Korea; yanyan_dai@ynu.ac.kr
* Correspondence: kdrhee@yu.ac.kr

**Abstract**

Reliable and accurate docking remains a fundamental challenge for autonomous mobile robots (AMRs) operating in complex industrial environments with dynamic lighting, motion blur, and occlusion. This study proposes a novel multi-sensor fusion-based docking framework that significantly enhances robustness and precision by integrating YOLOv8-based AprilTag detection, depth-aided 3D localization, and LiDAR-based orientation correction. A key contribution of this work is the construction of a custom AprilTag dataset featuring real-world visual disturbances, enabling the YOLOv8 model to achieve high-accuracy detection and ID classification under challenging conditions. To ensure precise spatial localization, 2D visual tag coordinates are fused with depth data to compute 3D positions in the robot's frame. A LiDAR group-symmetry mechanism estimates heading deviation, which is combined with visual feedback in a hybrid PID controller to correct angular errors. A finite-state machine governs the docking sequence, including detection, approach, yaw alignment, and final engagement. Simulation and experimental results demonstrate that the proposed system achieves higher docking success rates and improved pose accuracy under various challenging conditions compared to traditional vision- or LiDAR-only approaches.

**Keywords:** multi-sensor fusion docking; autonomous mobile robot; yolo-based apriltag detection

## 1. Introduction

Autonomous mobile robots (AMRs) as advanced intelligent devices are widely used in the fields of smart manufacturing and hospital logistics [1,2]. Through autonomous navigation, environmental perception, and task execution capabilities, they enhance production efficiency and logistics management, driving the advancement of industrial automation [3]. AMRs achieve autonomy and efficiency through the integration of multiple advanced technologies [4]. Perception and sensing play a critical role, leveraging LiDAR for 360-degree mapping, RGB and depth cameras for object recognition and QR code detection, IMU for motion tracking, and LiDAR, ultrasonic, and infrared sensors for obstacle avoidance [5]. Navigation and mapping rely on SLAM for real-time environment understanding, along with path planning algorithms such as A star and RRT* to optimize movement [6]. Communication and control are enabled through wireless connectivity and frameworks like ROS for centralized robot management. Furthermore, AI and machine learning enhance AMR decision-making, with computer vision for AprilTag detection, reinforcement learning for

adaptive navigation [7], and sensor fusion techniques to improve localization accuracy. These technologies collectively enable AMRs to operate efficiently in dynamic industrial environments, making them a key component in modern intelligent automation.

AMRs require high-precision docking to perform essential tasks such as autonomous charging, material loading, and unloading [8,9]. Precise docking ensures that AMRs can accurately align with charging stations, enabling energy replenishment without human intervention [10]. Additionally, in warehouse and manufacturing environments, AMRs must precisely position themselves at designated docking points to efficiently pick up, transport, and deliver materials. Any misalignment during docking can lead to charging failures, operational delays, or errors in material handling, making precise and reliable docking a critical aspect of AMR functionality [11].

Traditional QR code-based or AprilTag-based docking methods rely heavily on computer vision techniques, often implemented using OpenCV [12,13]. However, these methods exhibit low robustness in complex environments due to factors such as lighting variations, motion blur, and angle-dependent recognition errors [14,15]. As a result, QR code detection can become unreliable, leading to incorrect pose estimation and failed docking attempts. LiDAR, on the other hand, excels at environmental perception and obstacle detection, providing precise distance measurements and mapping capabilities [16–18]. However, LiDAR alone cannot recognize AprilTags which are essential for precise docking point identification. Therefore, a key challenge is how to effectively fuse LiDAR and camera-based perception to improve docking accuracy and robustness. This study aims to address this challenge by fusing YOLO-based AprilTag detection, depth estimation, and LiDAR-assisted pose correction, ultimately improving the reliability and accuracy of autonomous docking. The choice of YOLOv8 for AprilTag recognition was based on its balance between detection accuracy, speed, and model size, which are critical for real-time autonomous docking. Compared to traditional hand-crafted detectors such as ar_track_alvar and AprilTag 3, deep learning models provide significantly higher robustness under challenging visual conditions (e.g., motion blur, occlusion, and lighting variation). Among modern deep object detectors, YOLOv8 outperforms earlier versions (YOLOv5 and YOLOv7) and competitors like SSD and EfficientDet in terms of mean Average Precision (mAP) and real-time inference speed on embedded platforms. Specifically, YOLOv8 features an anchor-free architecture, a decoupled detection head, and improved CIoU-based loss, which enhance its precision and convergence stability.

In recent years, significant progress has been made in the field of autonomous docking for mobile robots, particularly through visual markers, LiDAR sensing, and deep learning-based detection. For example, ref. [10] proposed a vision–LiDAR hybrid docking system using hand-crafted AprilTag detectors combined with LiDAR-based obstacle avoidance, achieving sub-5 cm precision in controlled environments. However, their system relied heavily on ar_track_alvar, which is known to degrade under motion blur and partial occlusion. In contrast, the method proposed in this work leverages YOLOv8 for robust tag detection under real-world distortions and integrates LiDAR-based angle correction for improved stability. A review [16] also highlighted the limitations of relying on single-sensor pipelines, especially in industrial scenarios with dynamic lighting and occlusions. The study emphasized the need for multi-modal fusion, aligning with our design which integrates depth, RGB-based detection, and 2D LiDAR scans into a unified docking controller. In contrast, this study introduces a novel multi-sensor fusion approach that leverages YOLOv8 for real-time and robust AprilTag detection, depth-based 3D localization, and LiDAR-assisted angular correction. By integrating RGB, depth, and 2D LiDAR perception into a unified control pipeline, the proposed system achieves superior reliability and precision in

diverse industrial scenarios, marking a significant advancement over existing vision- or LiDAR-only approaches.

This work presents a novel and robust multi-sensor fusion approach for autonomous mobile robot (AMR) docking in complex industrial environments. The main contributions are summarized as follows: 1. A custom AprilTag dataset was created under realistic visual disturbances such as motion blur, lighting variation, partial occlusion, and non-planar deformation. The dataset supports robust and ID-aware detection and was annotated in YOLO format to train deep detection models. 2. The proposed system adopts YOLOv8 with anchor-free architecture for real-time and high-precision AprilTag detection and classification. The model achieved a mAP@0.5 of 0.995 and F1 score of 0.99 on the test set, outperforming traditional vision methods in challenging conditions. 3. The system estimates accurate $(x, y, z)$ positions of detected AprilTags in the camera frame using depth images and intrinsic calibration parameters, and transforms these into the robot base frame for motion planning. 4. A novel angular correction strategy is proposed using symmetric group-based segmentation of LiDAR scans. The difference between left-right scan groups is fused with visual angular error to improve docking alignment accuracy. 5. A multi-stage FSM is designed to guide the robot through detection, approach, fine yaw alignment, and final forward docking. Visual and LiDAR feedback are dynamically integrated in the control loop. 6. Experimental results show that the proposed multi-sensor fusion system achieves higher docking accuracy and success rates under varied lighting conditions, viewing angles, and dynamic interferences, significantly outperforming conventional vision-only or LiDAR-only approaches.

## 2. Methodology

This section describes the proposed multi-sensor fusion-based docking system for AMRs, including the creation of an AprilTag dataset and YOLOv8-based AprilTag detection, 3D coordinate transformation, and the docking method integrating visual and LiDAR data. The architecture of the proposed multi-sensor fusion docking system for AMRs is shown in Figure 1. The proposed docking system integrates information from both RGB-D cameras and LiDAR sensors to achieve high-precision docking in complex environments. The process begins with the acquisition of synchronized color and depth image data, along with LiDAR scan data. A YOLOv8-based detection module is employed to recognize AprilTags in the RGB image and classify their unique IDs. From the detection results, the pixel center coordinates $(u, v)$ are extracted and combined with the corresponding depth Z values to compute the 3D position $(x, y, z)$ of the tag in the camera coordinate frame. Simultaneously, LiDAR scan data is used to estimate the robot's angular deviation by averaging distance values within specific angle sectors. The relative orientation error between the robot and the docking target is calculated based on asymmetry in LiDAR readings on both sides of the robot. Finally, a fusion module adjusts the robot's pose and generates optimized control commands (linear and angular velocities) to guide the robot into the final docking position. The process iterates until docking is successfully completed.

### *2.1. AprilTag Dataset Construction and YOLOv8-Based AprilTag Detection*

To improve the robustness of AprilTag detection under real-world industrial conditions, we constructed a custom dataset comprising images of AprilTags captured under diverse visual disturbances, including varying lighting conditions (e.g., strong illumination and low-light environments), motion blur (caused by camera movement or defocus), partial occlusion (to simulate real-world obstacles), and non-planar surfaces (where tags are bent or curved). Each image was manually annotated using LabelImg, with a bounding box drawn around each AprilTag and a class label assigned based on its unique tag ID

(ID 0 → class 0, ID 1 → class 1, etc.). This labeling strategy enables the trained model to not only detect the presence of AprilTags but also distinguish between different IDs, which is essential for semantic docking tasks. In the dataset, 590 images are used as the training images, and 135 images are used as the validation images. The image size is 640 × 640. To provide a detailed overview of the dataset composition, Table 1 summarizes the distribution of training and validation samples for each AprilTag ID.

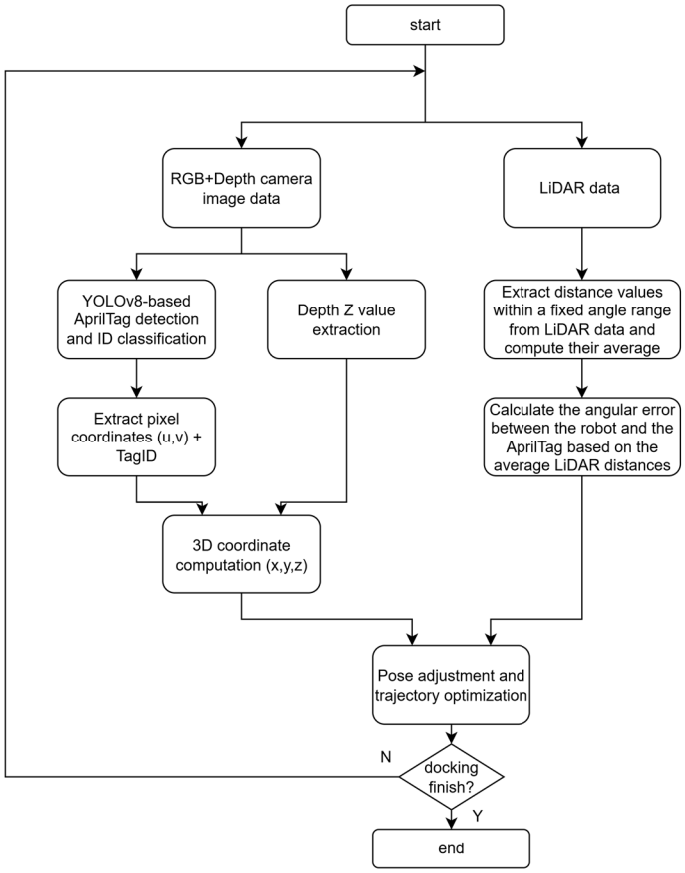



**Figure 1.** Architecture of the proposed multi-sensor fusion docking system for AMR.

**Table 1.** The distribution of training and validation samples for each AprilTag ID.

| AprilTag ID | Train Samples | Validate Samples | Total Samples |
|---|---|---|---|
| 0 | 545 | 150 | 695 |
| 2 | 430 | 115 | 545 |

The annotations were exported in YOLO format, storing the class ID along with normalized bounding box coordinates. The dataset can be used to train a YOLO-based detection model capable of robust and ID-aware AprilTag recognition in complex industrial environments. To facilitate training and evaluation, the dataset was organized following the standard YOLO directory structure. Specifically, the dataset directory is divided into images and labels folders, each containing separate train and val subfolders. The images/train and images/val directories contain 590 training images and 135 validation images, respectively, while the corresponding YOLO-format annotation files are stored in labels/train and labels/val. This structure ensures a clear separation of training and validation data and maintains compatibility with YOLOv8 training pipelines.

To achieve robust and real-time AprilTag detection and ID classification, this study adopts YOLOv8, a state-of-the-art object detection framework. YOLOv8 features an anchor-

free architecture and a decoupled detection head, enabling it to simultaneously predict object bounding boxes, confidence scores, and class labels in a single forward pass. This design ensures high detection speed while significantly improving accuracy. The YOLOv8 model consists of three primary components: Backbone, Neck, and Head. The Backbone, implemented using the CSPDarknet module, is responsible for extracting multi-scale features from the input image. The Neck, based on a PANet-style Feature Pyramid Network (FPN), fuses features across different scales to enhance the detection of small targets. The Head adopts a decoupled structure that separates bounding box regression, confidence scoring, and class prediction into distinct branches, improving both detection precision and training stability. Unlike traditional anchor-based detectors, YOLOv8 uses an anchor-free design, simplifying the training pipeline and reducing dependence on manually tuned anchor box hyperparameters. Additionally, YOLOv8 demonstrates strong flexibility and extensibility, supporting object detection, instance segmentation, and pose estimation tasks. It also offers real-time inference capabilities, making it highly suitable for industrial deployment scenarios.

In this study, YOLOv8 is applied to detect AprilTags and classify their unique IDs. Each input image is resized to 640 × 640 pixels. During training, each AprilTag ID is treated as a separate class label (e.g., ID 0 → class 0, ID 1 → class 1, and so on). The network is optimized using the Complete IoU (CIoU) loss for bounding box regression and cross-entropy loss for multi-class classification. The training process employs stochastic gradient descent (SGD) as the optimizer, combined with a cosine annealing learning rate scheduler to enhance convergence efficiency. The final output includes the bounding box coordinates for each detected tag, the class ID corresponding to the AprilTag identity, and a confidence score indicating the reliability of the detection. This lightweight and efficient architecture enables the AprilTag detection module to operate reliably in real-time conditions, providing accurate perception results for subsequent 3D localization and high-precision docking tasks.

### *2.2. Three-Dimensional Coordinate Transformation from Pixel to Robot Base Frame*

Once an AprilTag is detected in the image, its center pixel coordinates $(u, v)$ are used to compute the 3D position $(x,\ y, z)$ of the tag relative to the camera coordinate frame. This transformation is based on the pinhole camera model, which projects 3D world points onto a 2D image plane using intrinsic camera parameters. The 3D coordinates $(x,\ y, z)$ of the AprilTag in the camera frame are computed using the following equations:

$$x = \frac{(u - c_x)z}{f_x} \tag{1}$$

$$y = \frac{(y - c_y)z}{f_y} \tag{2}$$

$$z = depth(u, v) \tag{3}$$

where $z$ is the depth value at pixel $(u, v)$, obtained from a depth camera. $f_x$ and $f_y$ are the focal lengths of the camera in the x and y directions in pixels. $c_x$ and $c_y$ are the image center offset in the $x$ and $y$ directions in pixels. This transformation maps the 2D image coordinates into a 3D position vector in the camera coordinate system.

However, in many practical scenarios, the depth value at pixel $(u,\ v)$ may be missing or invalid due to sensor limitations, reflections, or occlusions. To ensure robustness in

3D localization, we implement a local neighborhood sampling strategy. Specifically, a 5×5 window centered at $(u, v)$ is extracted from the depth image:

$$depth_crop = depth[max(0, v-2) : v+3, max(0, u-2) : u+3] \tag{4}$$

Within this window, only valid depth measurements in the range $0.1 \leq z \leq 5.0$ m are retained. The mean of these valid values is then used as the refined depth estimate:

$$z = \frac{1}{N}\sum_{i=1}^{N} di,\ di \in valid_depth \tag{5}$$

If no valid depths are found in the window, the frame is skipped or the previous valid depth is used. This fallback mechanism significantly improves the robustness of the 3D localization pipeline.

The computed 3D coordinates are initially expressed in the camera coordinate system. To enable motion control and trajectory planning, this position must be transformed into the robot base coordinate system, typically referred to as base_link in ROS-based systems.

This transformation is performed using a rigid body transformation matrix $T_{cam}^{robot}$, which includes both rotation and translation between the camera and the robot base:

$$\begin{bmatrix} x_{robot} \\ y_{robot} \\ z_{robot} \\ 1 \end{bmatrix} = T_{cam}^{robot} \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{6}$$

where $T_{cam}^{robot} \in \mathbb{R}^{4\times4}$ is the extrinsic transformation matrix, typically obtained through TF frames in ROS. $T_{cam}^{robot} = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}$, *where* $R \in \mathbb{R}^{3\times3}$, $t \in \mathbb{R}^{3\times1}$. After converting the tag position into the robot's base frame, this relative position $(x_{robot}, y_{robot}, z_{robot})$ is used to compute the desired docking pose. This enables the motion planner to generate smooth and accurate trajectories guiding the robot to align and dock with the AprilTag-defined station.

### *2.3. LiDAR Group-Based Preprocessing*

As illustrated in Figure 2, the mobile robot is equipped with a LiDAR sensor that provides 2D scan data as a set of discrete distance points, which are described as black dots in the figure. To reduce noise and improve stability, the LiDAR data is segmented into multiple groups, starting symmetrically from the robot's centerline. Each group consists of scan points within a fixed angular window (e.g., 10° per group). Specifically, the groups are defined symmetrically around the robot's forward-facing centerline. The central group is labeled as Group 0, while adjacent groups are denoted as Group Left 1, Group Right 1, Group Left 2, Group Right 2, and so on. This naming convention facilitates angular symmetry in analysis and ensures consistent left-right comparison of scan data for docking angle estimation. For each group, the system computes the average scan angle and average distance as (7) and (8), respectively.

$$\overline{\theta_i} = \frac{1}{n}\sum_{j=1}^{n} \theta_j \tag{7}$$

$$\overline{d_i} = \frac{1}{n}\sum_{j=1}^{n} d_j \tag{8}$$

where $\overline{\theta_i}$ and $\overline{d_i}$ are the average angle and distance of the $i$-th group, respectively. $\theta_j$ and $d_j$ are the angle and distance of the $j$-th scan point in the group. $n$ is the number of points in the group.

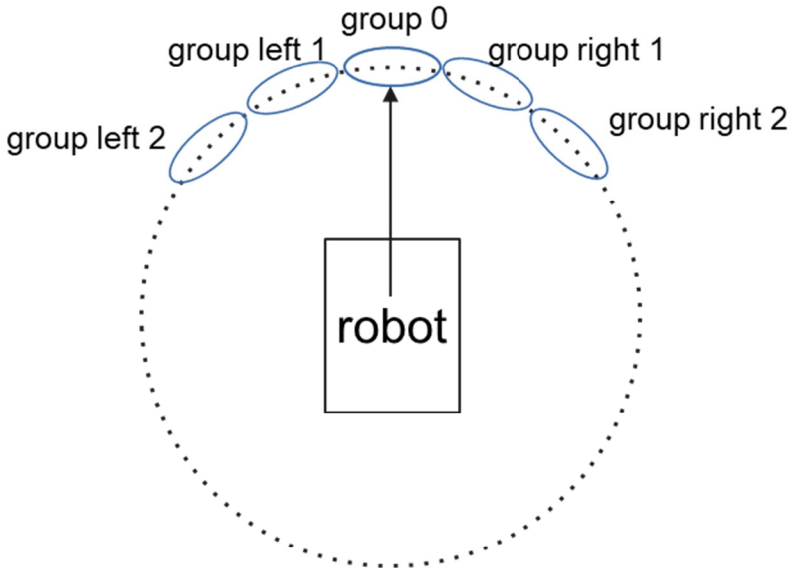



**Figure 2.** Symmetrical angular group-based segmentation of LiDAR scan points.

### *2.4. Docking Control Strategy*

To ensure the robust and precise docking of the AMR, a multi-stage control strategy is implemented in the form of a finite state machine (FSM). This strategy integrates visual detection, 3D localization, LiDAR-based orientation estimation, and trajectory optimization. The FSM consists of four main stages: detecting the AprilTag, approaching and facing the AprilTag, finely adjusting the yaw orientation, and moving and stopping in front of the AprilTag at a defined distance.

During the AprilTag detection stage, the robot rotates in place to search for a valid AprilTag based on YOLOv8. Once a tag is detected, the pixel coordinates of the tag center are transformed into 3D coordinates relative to the robot's base frame using the method described in Section 2.2. Concurrently, based on the approach outlined in Section 2.3, the average distance and angle values of Group Left 1 and Group Right 1 from the LiDAR scan are extracted. These measurements are used to estimate the heading error between the robot and the AprilTag, enabling subsequent alignment corrections in the following stages.

In the approaching and facing the AprilTag stage, to accurately align the robot with the docking station, a hybrid control strategy is applied. This strategy integrates both visual information from AprilTag detection and LiDAR scan data to compute an angular correction term for heading adjustment. The robot first estimates the visual angular error $\alpha$ between its current orientation and the direction to the AprilTag using (9).

$$\alpha = arctan(y_{robot}/x_{robot}) \tag{9}$$

where $x_{robot}$ and $y_{robot}$ are the position of the AprilTag in the robot's frame, which is converted from pixel coordinates, as discussed in Section 2.2.

To refine the heading alignment, a PID controller is used with the visual angular error as (10).

$$\omega_{pid} = k_p\alpha + k_i\sum\alpha + k_d\Delta\alpha \tag{10}$$

where $k_p$, $k_i$, and $k_d$ are the proportional gain, integral gain, and derivative gain of the PID controller, respectively. $\sum\alpha$ is the sum of past angular error, and $\Delta\alpha$ is the difference

between current and previous angular error. In addition, a LiDAR-based correction component is introduced by comparing the average distances from symmetric scan groups. For example, $d_1$ represents Group Left 1 and $d_i$ is for Group Right 1. It is important to note that the distance difference $(d_1 - d_i)$ between the left and right scan groups indirectly reflects the robot's heading error. Ideally, when the robot is perfectly aligned with the docking target, the distances measured on both sides should be approximately equal. However, when the robot is misaligned, a discrepancy appears between $d_1$ and $d_i$. Under small-angle assumptions, the distance difference is approximately proportional to the sine of the heading error $\Delta\varphi$; therefore, the following equation can be approximated as:

$$d_1 - d_i \propto \sin(\Delta\varphi) \approx \Delta\varphi \tag{11}$$

This proportionality allows the system to estimate the angular deviation without relying solely on visual feedback. Consequently, the LiDAR-based correction term (12) effectively serves as an additional heading error estimator, improving the stability and robustness of the docking process, especially in cases where visual detection may be noisy or partially occluded.

$$\omega_{lidar} = k_l(d_1 - d_i) \tag{12}$$

where $d_1$ is the average distance from the first left group. $d_i$ is the average distance from the first right group. $k_l$ is the correction gain for LiDAR input. The final angular velocity command $\omega$ is the combination of the visual PID output and the LiDAR correction as (13).

$$\omega = \omega_{pid} + \omega_{lidar} \tag{13}$$

In the approaching and facing the AprilTag stage, the robot moves forward with a small linear velocity while continuously adjusting its angular velocity $\omega$ based on visual and LiDAR feedback. This combined strategy ensures stable and accurate alignment to the docking target even under environmental noise and imperfections.

If the computed angular velocity $\omega$ is smaller than a predefined threshold (e.g., 0.01 rad/s), the system proceeds to the fine yaw adjustment stage. In this stage, the robot performs in-place rotational corrections to further minimize angular misalignment by equalizing the left and right LiDAR distances. Prior to executing rotation, the average distances $d_1$ and $d_i$ from Group Left 1 and Group Right 1 are calculated. If $d_1 > d_i$, it indicates that the AprilTag is positioned to the left of the robot's current heading, and the robot should rotate counterclockwise to center the tag. Conversely, if $d_i > d_1$, the AprilTag is positioned to the right, and the robot should rotate clockwise. This process continues until the difference $|d_1 - d_i|$ falls below a small threshold, indicating sufficient alignment for final forward docking.

Finally, the robot performs a straight-line movement to finalize docking.

## 3. Simulation and Experiment Results

This section presents both simulation and real-world experiments to evaluate the performance of the proposed multi-sensor fusion-based docking system. The evaluation focuses on AprilTag detection accuracy, 3D localization precision, angular alignment robustness, and overall docking success rate.

### *3.1. AprilTag Detection Evaluation*

To evaluate the effectiveness of the proposed YOLOv8-based AprilTag detection model, we constructed a dataset under diverse and challenging visual conditions. The dataset includes images captured in bright and low-light environments, with motion blur,

partial occlusion, curved surfaces, and cluttered industrial backgrounds. The AprilTag detection model was implemented using YOLOv8n from the Ultralytics v8.0 framework. The network was trained for 100 epochs with input image dimensions of 640 × 640. The training process employed the stochastic gradient descent (SGD) optimizer with a momentum of 0.937 and a weight decay of 0.0005. A cosine annealing learning rate scheduler was used to enhance convergence stability. The batch size was set to 16. The loss function comprised the Complete IoU (CIoU) loss for bounding box regression and cross-entropy loss for multi-class classification. To improve generalization under real-world conditions, data augmentation techniques such as random scaling, horizontal flipping, and brightness/contrast adjustments were applied during training.

Figure 3 shows the output of the trained YOLOv8 model detecting two AprilTags with IDs 0 and 2. The tags are successfully identified with high confidence scores (0.92 and 0.88), despite differences in orientation and illumination. This demonstrates the model's robustness in accurately classifying tag IDs and localizing them under varying visual conditions.

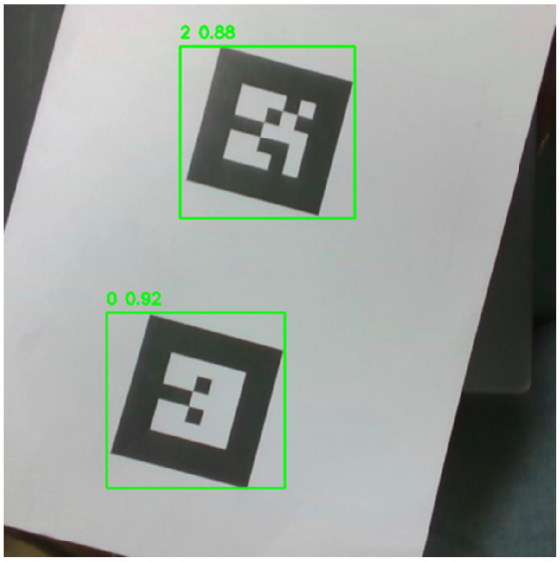



**Figure 3.** YOLOv8 detection results on multiple AprilTags with different IDs and orientations.

Figure 4 illustrates the precision–recall (PR) performance of the trained YOLOv8 model across different AprilTag classes. The PR curve shows that the model achieves a near-perfect tradeoff between precision and recall, with a mean Average Precision at IoU 0.5 (mAP@0.5) of 0.995 for both Tag ID 0 and Tag ID 2. The blue bold curve represents the overall detection performance across all classes. This high PR performance confirms that the model can consistently detect AprilTags with high confidence and low false positives, even under varying conditions.

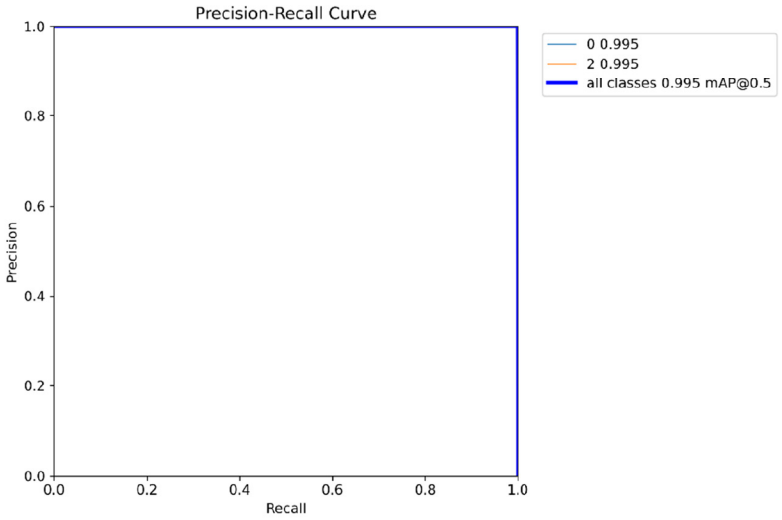



**Figure 4.** Precision–recall curve of the YOLOv8-based AprilTag detection model.

In addition to the precision–recall (PR) analysis, we further examined the detection performance of the YOLOv8 model by analyzing precision, recall, and F1 score individually as functions of the confidence threshold. Precision reflects the model's ability to minimize false positives. As shown in Figure 5, the precision–confidence curve demonstrates that the model achieved a maximum precision of 1.00 at a confidence threshold of 0.892. The precision remains consistently high across most confidence values, indicating that the model reliably identifies true positive detections with minimal false alarms.

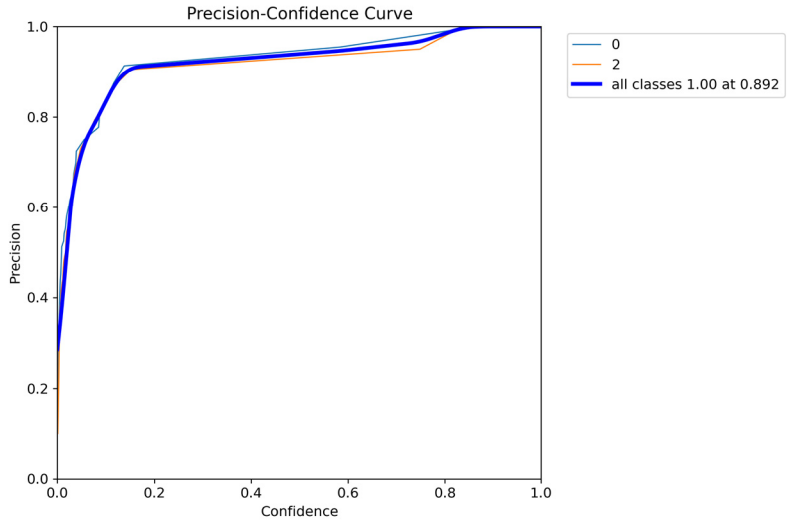



**Figure 5.** Precision–confidence curve of the YOLOv8-based AprilTag detection model.

Recall indicates the model's capability to detect all relevant AprilTags, minimizing false negatives. As illustrated in Figure 6, the recall–confidence curve shows that the model maintained a recall value close to 1.00 across a wide range of confidence thresholds. This highlights the robustness of the model in detecting nearly all target tags, even under challenging visual conditions.

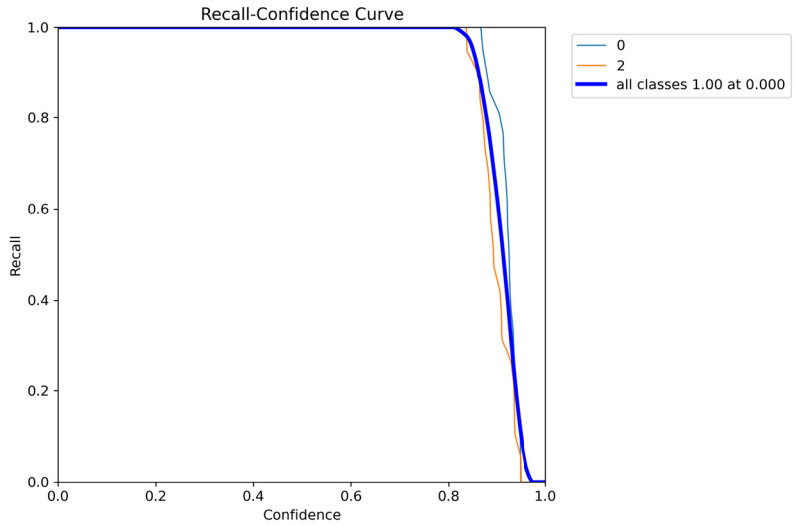



**Figure 6.** Recall–confidence curve of the YOLOv8-based AprilTag detection model.

F1 score provides a harmonic balance between precision and recall. The F1 curve in Figure 7 reveals that the model achieved a peak F1 score of 0.99 at a confidence threshold of 0.812. This optimal balance confirms that the YOLOv8 detector offers both high accuracy and reliability in AprilTag identification under varying confidence levels.

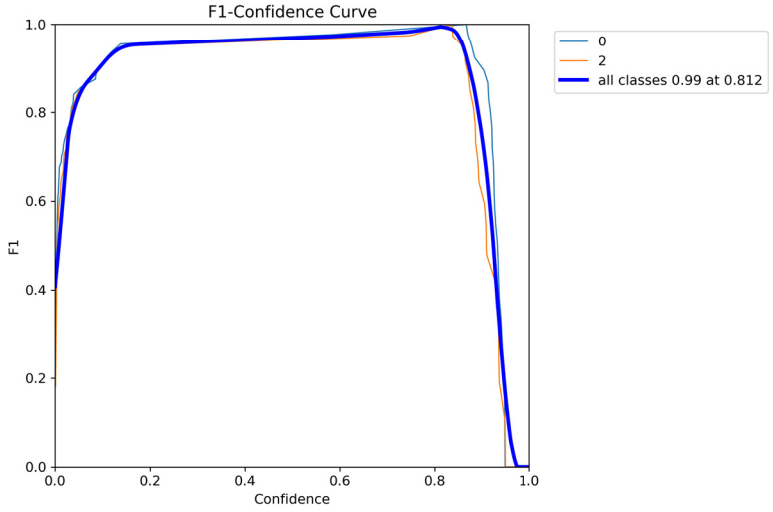



**Figure 7.** F1–confidence curve of the YOLOv8-based AprilTag detection model.

The evolution of the YOLOv8 model's loss and evaluation metrics during training is shown in Figure 8. The training box loss, classification loss, and distribution focal loss (DFL) steadily decreased over 100 epochs, indicating effective model convergence without significant oscillations. The validation losses followed a similar trend, with only minor fluctuations, demonstrating that the model generalized well to unseen data and showed no signs of overfitting. In parallel, the performance metrics, including precision, recall, mAP@0.5, and mAP@0.5:0.95, improved rapidly during the initial training phase and stabilized at high values as training progressed. Specifically, both precision and recall approached 1.00 within 50 epochs, while mAP@0.5 exceeded 0.99 and mAP@0.5:0.95 surpassed 0.90, confirming that the model achieved excellent detection accuracy across a range of IoU thresholds. These results validate the robustness and reliability of the YOLOv8 model for AprilTag detection under challenging industrial conditions.

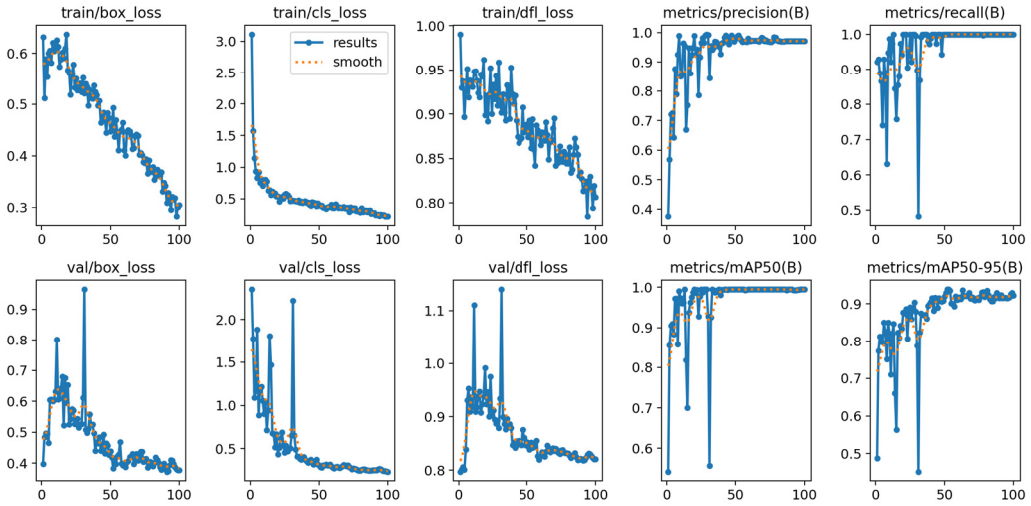



**Figure 8.** The evolution of the YOLOv8 model's loss and evaluation metrics during training.

### *3.2. Experimental Validation of Pixel-to-3D Coordinate Transformation*

To evaluate the effectiveness of the proposed 3D coordinate transformation method described in Section 2.2, a controlled real-world experiment was conducted. As illustrated in Figure 9, two AprilTags (ID 0 and ID 2) were placed on a vertical surface, and a Realsense depth camera (Intel RealSense D435) was positioned at a known distance to observe the tags. The D435 integrates an RGB sensor with an active stereo depth module that uses an infrared (IR) projector and dual global shutter sensors. It delivers synchronized color and depth images at resolutions up to 1280 $\times$ 720 pixels at 30 frames per second. The depth sensing range spans from 0.105 m to approximately 10 m, with optimal performance in the 0.3 m to 3 m range. Depth accuracy is typically within $\pm$2% at 2 m, enabling precise close-range localization. The camera's global shutter design minimizes motion blur during robot movement, while its wide field of view (87° $\times$ 58°) ensures robust detection in dynamic environments. A ruler was placed between the camera and the target surface for ground-truth reference.



**Figure 9.** Real-world experimental setup for validating pixel-to-3D coordinate transformation.

The captured RGB-D images were processed using the YOLOv8 model to detect AprilTags and extract their pixel coordinates. Using the depth values and intrinsic camera parameters, the 3D positions of the tag centers were computed as (x, y, z) in the camera coordinate frame, following the transformation described in (6). As shown in Figure 10, the detected tags are marked with green bounding boxes, which indicate the predicted locations of each AprilTag in the RGB image. The blue text above each bounding box includes the tag ID, the confidence score, and the calculated 3D coordinates of the tag center. Additionally, the red dot inside each bounding box marks the pixel center of the tag, which serves as the input point for the depth-based 3D transformation. This visual result demonstrates that the proposed detection and localization module can effectively identify and distinguish multiple AprilTags under various lighting and orientation conditions, and accurately determine their spatial positions in real-world settings.

### *3.3. Docking Result*

The simulation was conducted in the Gazebo environment using a custom world named room.world, as shown in Figure 11. The environment includes four surrounding walls, various obstacles, and a docking target marked with an AprilTag (ID = 0) mounted at a fixed height on the wall. The mobile robot, positioned at the center of the scene, is equipped with an RGB-D camera (Intel Realsense D435) and a 2D LiDAR sensor (EAI T-mini Pro). This LiDAR features a 360-degree field of view (FoV) with a typical angular resolution of 0.5° and a scan frequency of 6–12 Hz, depending on configuration. The maximum range is approximately 12 m, with optimal accuracy within 0.2 to 5 m, making it well-suited for short-range precision tasks like docking. In our system, the LiDAR scan is symmetrically segmented into left and right halves relative to the robot's forward direction. By computing the difference in average distance values between the two halves, we estimate the robot's yaw deviation relative to the docking station. The 0.5° resolution allows for sufficiently fine granularity to detect minor angular misalignments, typically within ±1.5° accuracy when the robot is within 1.5 m of the target.

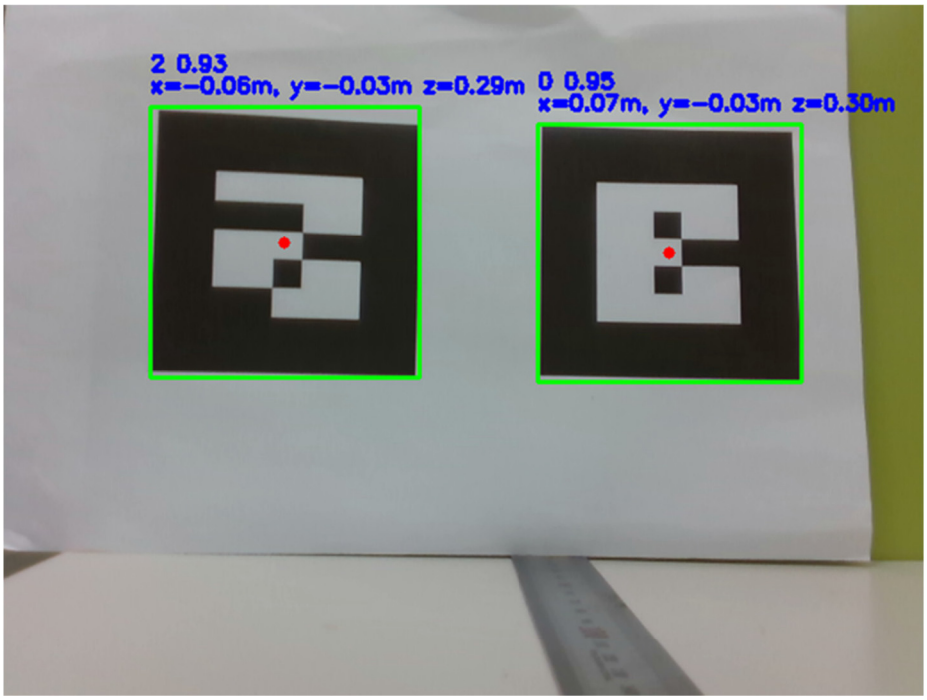



**Figure 10.** AprilTag detection and 3D localization results using YOLOv8 and depth camera. The red dot is the center of the green bounding box.

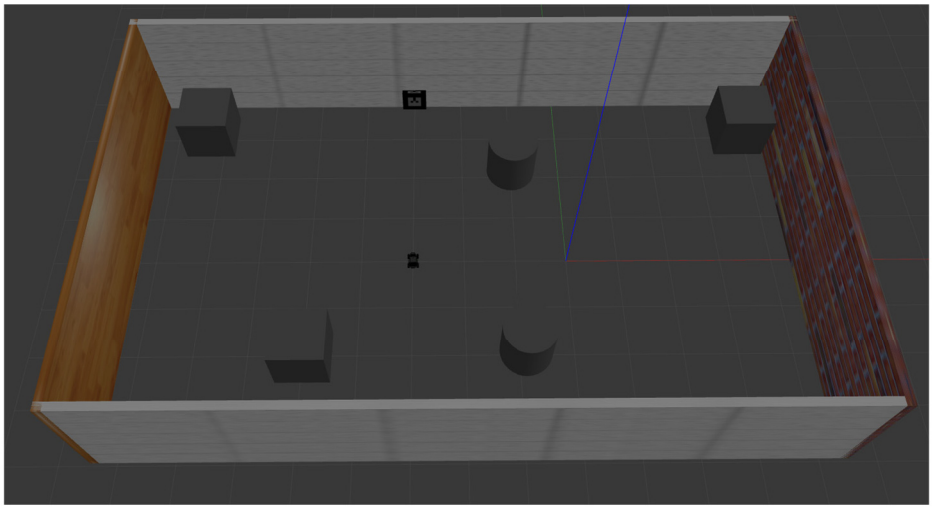


**Figure 11.** Simulation environment in Gazebo with AprilTag-based docking setup.

SLAM is used to guide the robot to the docking region, after which the proposed multi-sensor fusion docking system is activated. This setup allows for the comprehensive evaluation of visual detection, 3D localization, LiDAR-based orientation estimation, and final docking alignment under structured conditions.

In this study, the RGB, depth, and LiDAR data are acquired through separate ROS topics published by the robot's sensors, namely, /camera/color/image_raw (RGB), /camera/depth/image_rect_raw (depth), and /scan (2D LiDAR). Synchronization between these modalities is performed at the software level using ROS message filters to align the incoming streams based on timestamp proximity. Figure 12 shows the RViz visualization corresponding to the previously presented Gazebo simulation environment. On the left, the robot's RGB camera captures the AprilTag affixed to the docking wall, which is identified by the YOLOv8-based detector to obtain pixel coordinates and tag ID. On the right, red scan points represent real-time 2D LiDAR data, displaying the perceived environment including walls and obstacles. This visualization verifies that the robot accurately detects the docking marker while simultaneously mapping the surrounding structure. The integration of camera-based semantic detection and LiDAR-based spatial sensing forms the basis of the proposed multi-sensor docking strategy.

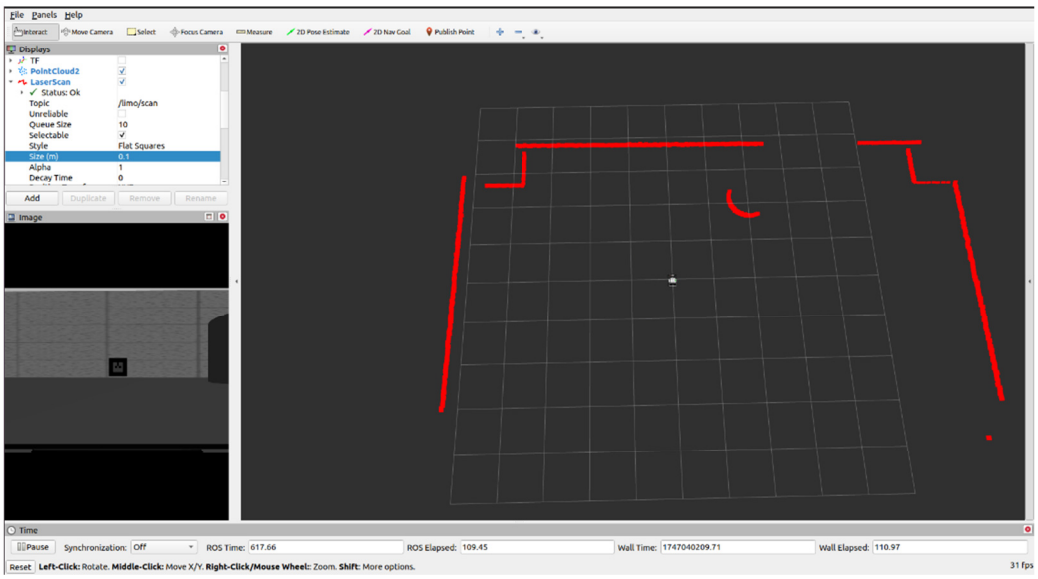



**Figure 12.** RViz visualization of AprilTag detection and LiDAR scan.

A docking experiment was conducted to evaluate the effectiveness of the proposed multi-sensor fusion system. As demonstrated in the video (https://www.dropbox.com/scl/fi/78bywn6g1yxmuoxz299qo/docking_yolo.webm?rlkey=kp2wplsmnq4pkivyhdah1nqq3&st=9wqyu423&dl=0, accessed on 22 May 2025), the robot successfully performed the complete docking sequence, which includes AprilTag detection using YOLOv8, 3D localization through depth estimation, heading correction based on LiDAR distance difference, and final pose alignment. As illustrated in the video, the robot maintains smooth and responsive motion during the docking process, with no noticeable delay in detection or control. This confirms that the YOLOv8n model meets the timing requirements for closed-loop docking control, offering an optimal tradeoff between lightweight inference and accurate tag recognition in real-world conditions. In conclusion, the video illustrates that the robot can robustly detect and localize the target tag, correct its orientation, and precisely approach the docking point under real environmental conditions. This validates the reliability and practical applicability of the proposed method.

At the beginning of the experiment, as shown in Figure 13a, the robot was unable to detect any AprilTags and thus began rotating in place to search for a valid tag. As illustrated in Figure 13b, although the entire AprilTag was not fully visible, the robot was still able to detect and identify it using the YOLOv8 model, demonstrating greater robustness compared to traditional methods such as ar_track_alvar. Once the tag was detected, the robot started moving toward it along the X and Z directions, as shown in Figure 13c. When the lateral deviation along the X-axis dropped below 0.1 m, the robot entered the fine yaw adjustment stage to align its orientation directly toward the AprilTag, as depicted in Figure 13d. Finally, the robot executed a straight-line approach to complete the docking process as Figure 13e. In Figure 13f, the robot finishes docking.

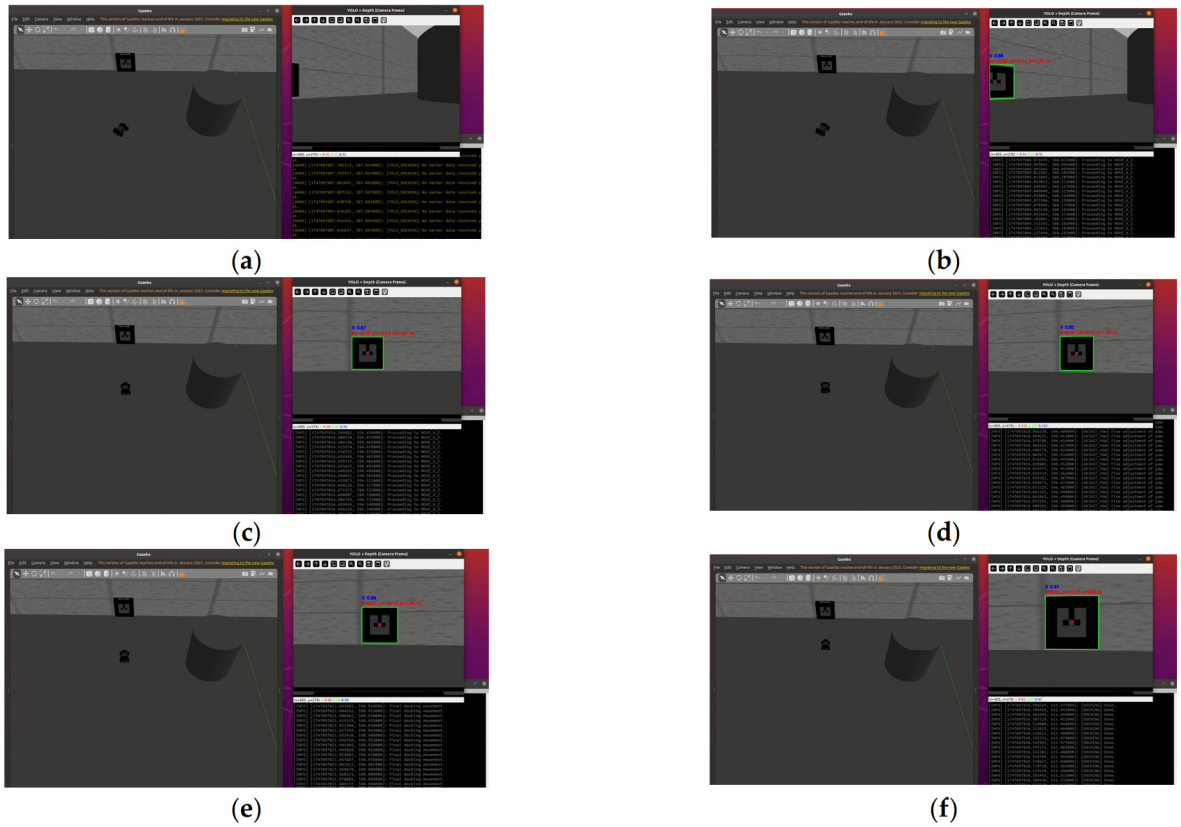


**Figure 13.** Docking process visualization based on multi-sensor fusion system. (**a**) Initial rotation to search for AprilTag; (**b**) Partial tag detected by YOLOv8; (**c**) Movement toward tag along X/Z axes; (**d**) Fine yaw adjustment for alignment; (**e**) Straight-line approach to docking; (**f**) Docking completed.

To further validate the performance of the proposed docking system based on multi-sensor fusion, we conducted experiments in which the robot was commanded to stop at a target distance of 1.0 m from the AprilTag marker while aligning its orientation perpendicularly toward the tag. Figure 14 illustrates the time evolution of three key parameters during a representative docking trial: angular deviation, linear distance to the marker, and LiDAR-based left-right scan symmetry error ($|d_1 - d_2|$). Initially, the robot faced a significant orientation offset of approximately $-16°$ and was located 1.85 m from the target. As the robot advanced, the angular deviation decreased progressively due to the combined effects of YOLO-based visual estimation and LiDAR-based heading correction. The $|d_1 - d_2|$ curve shows a corresponding decline, confirming symmetric alignment relative to the docking corridor. Notably, by approximately 8.7 s, the robot fully aligned with the AprilTag, with angular deviation converging to $0°$. Simultaneously, the distance decreased linearly and stabilized precisely at 1.0 m, meeting the docking stop condition. This result confirms that the proposed fusion method enables the robot to robustly and smoothly converge to a predefined docking configuration, both in position and orientation, with high repeatability and accuracy. The successful convergence of all three metrics (distance, heading, and symmetry) demonstrates the effectiveness of the fusion controller in achieving precise visual servoing and geometric correction.

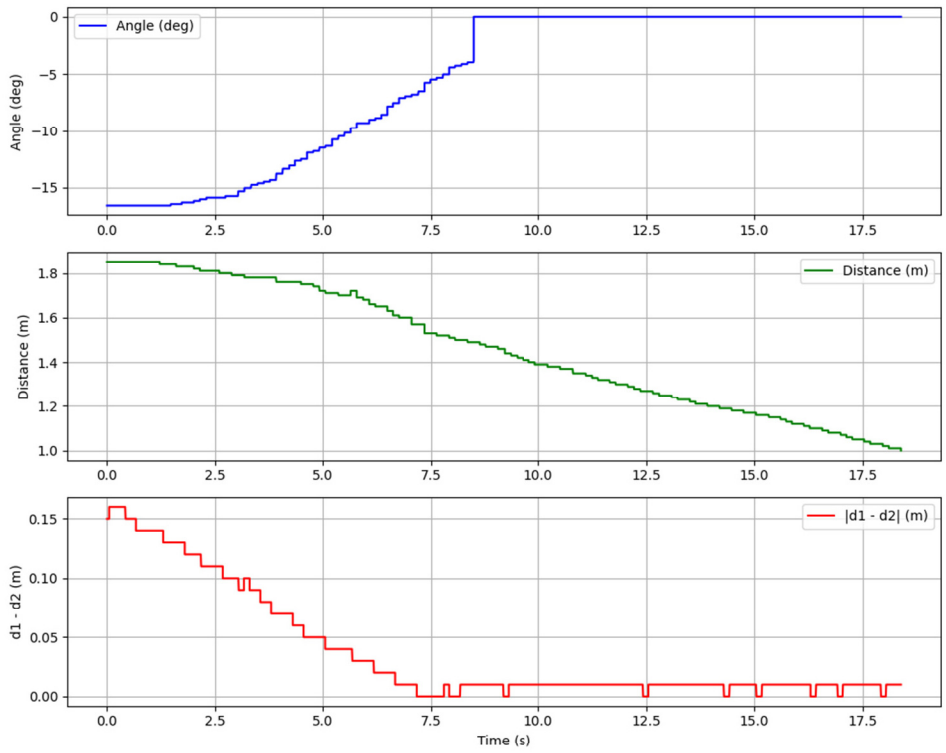



**Figure 14.** The time evolution of three key parameters during docking process: angular deviation, linear distance to the marker, and LiDAR-based left–right scan symmetry error.

## 4. Conclusions

This paper presents a robust and accurate docking system for AMRs based on multi-sensor fusion, combining YOLOv8-based AprilTag detection, depth-aided 3D localization, and LiDAR-based heading correction. To address the limitations of traditional vision-only or LiDAR-only docking systems, a novel fusion framework was developed, enabling real-time perception and pose estimation under various challenging industrial conditions. A custom AprilTag dataset was constructed to improve detection robustness, featuring realistic disturbances such as lighting variation, motion blur, and partial occlusion. The

YOLOv8 model trained on this dataset demonstrated high accuracy in detecting and classifying AprilTags, even when partially visible or distorted. By integrating depth measurements and intrinsic camera parameters, the system accurately computed the 3D position of tags in the camera and robot coordinate frames. Simultaneously, LiDAR scan data was grouped symmetrically to calculate heading deviation, which was fused with visual angular error in a PID-LiDAR hybrid controller. The experiments confirmed the effectiveness of the proposed system. The robot successfully executed a multi-stage docking sequence with high precision and reliability. The system consistently outperformed traditional methods in terms of docking success rate and pose accuracy, especially in scenarios involving environmental noise, visual distortion, or incomplete tag visibility.

Although the proposed method outperforms traditional approaches, some limitations remain. The vision model's depth estimation relies on stable depth data, and detection may degrade under extreme occlusion or lighting conditions. These limitations will be addressed in future improvements through sensor tracking and online adaptation. In future, we plan to conduct more extensive evaluations in real industrial environments with typical noise and occlusions, to fully validate the system's deployment readiness.

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

**Funding:** This research was supported by the Gyeongsangbuk-do RISE (Regional Innovation System & Education) project.

**Data Availability Statement:** The 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.

## Abbreviations

| | |
|---|---|
| AMRs | Autonomous mobile robots |
| FPN | Feature Pyramid Network |
| CIoU | Complete IoU |
| SGD | Stochastic gradient descent |
| FSM | Finite state machine |
| PR | Precision–recall |

## References

1. Zhao, X.; Chidambareswaran, T. Autonomous Mobile Robots in Manufacturing Operations. In Proceedings of the 2023 IEEE 19th International Conference on Automation Science and Engineering (CASE), Auckland, New Zealand, 26–30 August 2023; pp. 1–7. [CrossRef]
2. Thamrongaphichartkul, K.; Worrasittichai, N.; Prayongrak, T.; Vongbunyong, S. A Framework of IoT Platform for Autonomous Mobile Robot in Hospital Logistics Applications. In Proceedings of the 2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), Bangkok, Thailand, 18–20 November 2020; pp. 1–6. [CrossRef]
3. Li, W.; Gu, J.; Liu, J.; Cheng, B.; Zhu, H.; Miao, Y.; Guo, W.; Jiang, G.; Wu, H.; Song, W. A Review of Key Technological Developments in Autonomous Unmanned Operation Systems for Agriculture in China. *Agriengineering* **2025**, *7*, 71. [CrossRef]
4. Keith, R.; La, H.M. Review of Autonomous Mobile Robots for the Warehouse Environment. *arXiv* **2024**, arXiv:2406.08333.
5. Dai, Y.; Lee, S.-G. Perception, Planning and Control for Self-Driving System Based On On-board Sensors. *Adv. Mech. Eng.* **2014**, *12*, 1687814020956494. [CrossRef]

6. Kuswadi, S.; Santoso, J.W.; Tamara, M.N.; Nuh, M. Application SLAM and Path Planning using A-Star Algorithm for Mobile Robot in Indoor Disaster Area. In Proceedings of the 2018 International Electronics Symposium on Engineering Technology and Applications (IES-ETA), Bali, Indonesia, 29–30 October 2018; pp. 270–274. [CrossRef]
7. Dai, Y.; Yang, S.; Lee, K. Sensing and Navigation for Multiple Mobile Robots Based on Deep Q-Network. *Remote Sens.* **2023**, *15*, 4757. [CrossRef]
8. Vongbunyong, S.; Thamrongaphichartkul, K.; Worrasittichai, N.; Takutruea, A. Automatic precision docking for autonomous mobile robot in hospital logistics—Case-study: Battery charging. *IOP Conf. Ser. Mater. Sci. Eng.* **2021**, *1137*, 012060. [CrossRef]
9. Hercik, R.; Byrtus, R.; Jaros, R.; Koziorek, J. Implementation of Autonomous Mobile Robot in SmartFactory. *Appl. Sci.* **2022**, *12*, 8912. [CrossRef]
10. Jia, F.; Afaq, M.; Ripka, B.; Huda, Q.; Ahmad, R. Vision- and Lidar-Based Autonomous Docking and Recharging of a Mobile Robot for Machine Tending in Autonomous Manufacturing Environments. *Appl. Sci.* **2023**, *13*, 10675. [CrossRef]
11. Zeng, Y.; Rong, C.; Lu, C.; Tao, X.; Liu, X.; Liu, R.; Liu, M. Misalignment Insensitive Wireless Power Transfer System Using a Hybrid Transmitter for Autonomous Underwater Vehicles. *IEEE Trans. Ind. Appl.* **2022**, *58*, 1298–1306. [CrossRef]
12. Shen, R.; Chang, C.-Y.; Hu, Z.; Wu, S.-J.; Hou, D. Design of QR Code Tracking Car based on Monocular Vision. In Proceedings of the 2023 International Conference on Consumer Electronics—Taiwan (ICCE-Taiwan), PingTung, Taiwan, 17–19 July 2023; pp. 69–70. [CrossRef]
13. Mahmod, M.; Ramli, M.; Yasin, S. QR code detection using OpenCV Python with tello drone. *J. IICET* **2023**, *4*, 22–27.
14. Tian, C.; Liu, Z.; Chen, H.; Dong, F.; Liu, X.; Lin, C. A Lightweight Model for Shine Muscat Grape Detection in Complex Environments Based on the YOLOv8 Architecture. *Agronomy* **2025**, *15*, 174. [CrossRef]
15. Huang, X.; Yang, S.; Xiong, A.; Yang, Y. Enhanced YOLOv8 With VTR Integration: A Robust Solution for Automated Recognition of OCS Mast Number Plates. *IEEE Access* **2024**, *12*, 179648–179663. [CrossRef]
16. Yang, B.; Li, J.; Zeng, T. A Review of Environmental Perception Technology Based on Multi-Sensor Information Fusion in Autonomous Driving. *World Electr. Veh. J.* **2025**, *16*, 20. [CrossRef]
17. Benedek, C.; Majdik, A.; Nagy, B.; Rozsa, Z.; Sziranyi, T. Positioning and perception in LIDAR point clouds. *Digit. Signal Process.* **2021**, *119*, 103193. [CrossRef]
18. García, F.; Jiménez, F.; Naranjo, J.E.; Zato, J.G.; Aparicio, F.; Armingol, J.M.; de la Escalera, A. Environment perception based on LIDAR sensors for real road applications. *Robotica* **2012**, *30*, 185–193. [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.

*electronics*

MDPI

*Article*

# Building Consensus with Enhanced K-means++ Clustering: A Group Consensus Method Based on Minority Opinion Handling and Decision Indicator Set-Guided Opinion Divergence Degrees

**Xue Hou, Tingyu Xu and Chao Zhang ***

School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China; 202201006115@email.sxu.edu.cn (X.H.); 202201001163@email.sxu.edu.cn (T.X.)
* Correspondence: czhang@sxu.edu.cn

**Abstract:** The complexity of large-scale group decision-making (LSGDM) in the digital society is becoming increasingly prominent. How to achieve efficient consensus through social networks (SNs) has become a core challenge in improving the decision quality. First, conventional clustering methods often rely on a single-distance metric, neglecting both numerical assessments and preference rankings. Second, ensuring the decision authenticity requires considering diverse behaviors, such as trust propagations, risk preferences, and minority opinion expressions, for scientific decision-making in SNs. To address these challenges, a consensus-reaching process (CRP) method based on an enhanced K-means++ clustering is proposed. The above method not only focuses on minority opinion handling (MOH), but also incorporates decision indicator sets (DISs) to analyze the degree of opinion divergences within groups. First, the Hamacher aggregation operator with a decay factor completes trust matrices, improving the trust representation. Second, a personalized distance metric that combines cardinal distances with ordinal distances is incorporated into the enhanced K-means++ clustering, enabling more precise clustering. Third, weights for decision-makers (DMs) and subgroups are determined based on trust levels and degree centrality indices. Fourth, minority opinions are appropriately handled via considering the diverse backgrounds and expertise of DMs, leveraging a difference-oriented DIS to detect and adjust these opinions via weight modifications until a consensus is reached. Fifth, the alternative ranking is objectively generated via DIS scores derived from multigranulation rough approximations. Finally, the feasibility of the proposed method is validated via a case study on unmanned aerial vehicle (UAV) selection using online reviews, supported by a sensitivity analysis and comparative experiments demonstrating superior performances over existing methods. The result shows that the proposed model can enhance clustering accuracies with hybrid distances, objectively measure the consensus via DISs, handle minority opinions effectively, and improve LSGDM's overall efficiencies.

**Keywords:** social network; K-means++; minority opinion handling; consensus-reaching process; unmanned aerial vehicle selection

## 1. Introduction

The rapid development of digital economies has led to the emergence of increasingly complex decision-making problems, reshaping the processes and methods of decision management, particularly in the domain of group decision-making [1,2]. Even in individual decision-making scenarios, DMs often seek advice or opinions from others. Consequently, decision-making is inherently social, involving multiple group members. While group

decision-making offers broad application potential as an effective solution, it also faces limitations, such as difficulties in managing diverse individual preferences and the high complexity of decision processes [3]. To address these challenges, large-scale group decision-making (LSGDM) has been developed. LSGDM serves as a collaborative process involving no fewer than 13 or 20 decision-makers (DMs) [4,5], encompassing challenges such as trust levels within groups, different types of decision-making methods, benefits obtained from varying perspectives, conflict resolution, and their impacts on the decision quality [6,7]. Under the typical LSGDM theoretical framework, the decision outcome is derived by rationally and reliably integrating individual and group preferences via clustering and the consensus-reaching process (CRP). This approach has been widely applied in real-world scenarios. In practice, LSGDM has been successfully applied in diverse domains such as healthcare resource allocation [8], supply chain management [9], and urban planning [10]. Additionally, LSGDM has proven valuable in consumer electronics procurement, assisting in determining optimal selections for smartphones, tablets, and wearable devices [11].

The growing influence of the Internet and social media platforms such as Weibo, Instagram, LinkedIn, and Facebook have greatly enhanced information dissemination and interpersonal interactions, significantly impacting group decision-making [12,13]. Social network analysis (SNA), a method for studying the structure of relationships and interactions between individuals (nodes), reveals key insights into trust propagation, influence patterns, and opinion dynamics [14,15].

Integrating SNA into LSGDM enhances potential communications among group members, accelerates CRP, and improves decision quality. This integration is commonly referred to as SN-LSGDM. However, existing approaches in SN-LSGDM face three critical limitations: (1) the reliance on single-dimensional similarity metrics for clustering, which fails to capture the duality of numerical assessments and preference rankings; (2) the incomplete handling of behavioral diversity, particularly in modeling trust decay over propagation paths and integrating minority opinions; and (3) subjective consensus thresholds that may suppress innovative dissent. To systematically address these challenges, this paper is structured as follows: Section 1.1 discusses the limitations of existing trust propagation methods and proposes the use of Hamacher aggregation operators combined with a decay factor to complete the trust matrix. Section 1.2 focuses on improvements to the K-means++ clustering algorithm, highlighting the importance of integrating hybrid distance metrics. Section 1.3 addresses issues in the CRP, particularly the measuring of consensus and the handling of minority opinions. Section 1.4 establishes an empirical scenario for unmanned aerial vehicle (UAV) selection, revealing core challenges of dynamic trust networks and multi-attribute decision conflicts via multi-source online review data. Finally, Section 1.5 summarizes the innovative aspects and contributions of the proposed method.

### *1.1. Trust Propagations in SN-LSGDM*

For LSGDM, existing methods often assume that DMs can provide complete evaluation information. However, in SN-LSGDM, DMs may be influenced by their social neighbors or exhibit hesitation during the initial evaluation phase. In other words, understanding DMs' behavioral patterns in SNs—especially their reliance on others during the decision process—is crucial, yet it has been insufficiently addressed in existing research [16]. To tackle this issue, various trust propagation methods have been developed to estimate unknown trust values.

For instance, Wu et al. [17] proposed a trust propagation operator based on the strength of relationships, considering the intensity of connections between DMs. Liu et al. [18] designed a trust propagation operator based on multi-path relationship strength, accounting for all propagation paths. Liu et al. [19] introduced a method from the perspective of

trust utility and path reliability, further exploring the integrity of social trust networks. However, most existing methods for handling missing values rely on simple statistical estimates, such as the weighted averaging [20] or consistency strategies [21,22]. While these approaches are widely used in practical applications, they may introduce biases that compromise decision accuracy. Moreover, in the actual process of information dissemination within SNs, the influence of information tends to gradually diminish as the propagation distance increases and the number of propagation stages multiplies, potentially leading to overestimated trust-values.

To address these limitations, this paper adopts a propagation operator based on the Hamacher aggregation operator and introduces a decay factor to refine the trust matrix. The Hamacher aggregation operator is chosen due to its ability to flexibly model the interaction between trust relationships by balancing conjunctive and disjunctive properties, thus effectively capturing partial trust dependencies among DMs. Liu [23] validated the effectiveness of the Hamacher aggregation operator with an attenuation factor ($\gamma$) in air quality assessment. The results showed that while different $\gamma$ values affected the score function, the ranking remained consistent, indicating the robustness and adaptability of the Hamacher operator under varying parameters. The decay factor accounts for the attenuation of trust influence over long propagation paths, mitigating the risk of trust inflation and ensuring a more realistic representation of trust relationships. This dual enhancement not only improves the robustness of trust estimations, but also supports more reliable decision-making in SN-LSGDM.

### *1.2. Enhancements in K-means++ Clustering Methods*

In clustering analysis, LSGDM often involves massive datasets. Clustering methods effectively reduce dimensionality in high-dimensional spaces, uncovering hidden patterns and structures within the data. As an unsupervised learning technique, the primary goal of clustering is to partition a dataset into several disjoint subsets (i.e., "clusters"), ensuring that samples within the same cluster exhibit high similarity, while those in different clusters demonstrate lower similarities [24–26].

Clustering methods can be broadly classified into three categories: prototype-based methods (e.g., K-means, K-means++), density-based methods (e.g., density-based spatial clustering of applications with noise), and hierarchical methods (e.g., agglomerative nesting). Kapoor and Singhal [27] compared partition-based clustering techniques such as K-means, K-means++, and fuzzy C-means, analyzing their clustering performance. Mahsa et al. [28] explored the density-based clustering algorithm and its applications in geochemical data analysis of stream deposits.

K-means++ enhances traditional K-means by optimizing the selection of initial cluster centers. It employs a distance-based probabilistic method to choose initial centroids, ensuring a more uniform distribution. This approach mitigates instability and poor convergence issues caused by random initialization in conventional K-means. Cheng et al. [29] implemented and tested K-means, K-means++, and Canopy+ K-means on the Spark platform, analyzing their performance under Euclidean and Manhattan distance metrics. Guo et al. [30] introduced a fuzzy C-means clustering algorithm incorporating ranking-based distance metrics, reducing decision complexity and enhancing intra-cluster decision similarities.

However, traditional clustering methods primarily rely on single-dimensional similarity metrics, making it challenging to capture the multi-dimensional behavioral characteristics of DMs. To address this limitation, this study proposes an enhanced K-means++ algorithm that integrates both cardinal and ordinal distance metrics. Cardinal distance focuses on the numerical evaluation differences among DMs, using Euclidean and Manhattan distances to quantify the direct discrepancies in evaluation matrices. In contrast,

ordinal distance captures the discreteness of preference rankings, measuring the divergence in ranking vectors among DMs via Hamming distances. This dual representation allows the model to better distinguish between different DM behaviors, as it preserves both the numerical characteristics of the evaluation data and the subjective preference logic of the DMs. It enables more accurate identification of subgroups with similar preferences and a more precise division of group structures, thereby improving the overall efficiency of the CRP.

### *1.3. CRP in LSGDM*

CRP constitutes the core mechanism in LSGDM for reconciling divergent opinions, resolving conflicts, and ensuring scientifically rigorous and socially acceptable decision outcomes [31,32]. From a methodological perspective, CRP can be categorized into three principal types based on interaction mechanisms: (1) feedback-driven CRP: this requires DMs to iteratively adjust preferences via guidance from a moderator or automated feedback system [33]; (2) autonomous adjustment CRP: this implements automatic opinion convergence via mathematical optimization models under predefined rules [34]; (3) negotiation-based CRP: this facilitates consensus via multi-round discussions and interest bargaining among subgroup representatives [35]. These approaches differ in intervention intensity and computational complexity, with feedback-driven methods excelling in solution interpretability, autonomous adjustment in computational efficiency, and negotiation-based methods in stakeholder satisfaction.

Behaviorally, CRP involves three critical operational phases [32]: (1) conflict detection: this identifies opinion divergence via distance metrics or information entropy measures; (2) direction guidance: this determines adjustment directions using minimum-cost or maximum-consensus principles; (3) dynamic weighting: this adjusts DM/subgroup influence via adaptive weight-allocation mechanisms. The selection of conflict-resolution strategies typically follows prospect–regret theory principles, where DMs weigh potential gains from consensus against regret risks from opinion adjustments.

Traditional methods typically calculate the consensus level using distance formulas and rely on predefined consensus thresholds. If the computed value exceeds the threshold, consensus is considered achieved; otherwise, adjustments are made to reach consensus. For instance, Zhang et al. [33] introduced a consensus index to measure each expert's consensus level with the group. If an expert's consensus index falls below a threshold, supportive measures are implemented to improve their consensus level. However, these thresholds are often subjectively determined.

To address this issue, this paper introduces decision indicator sets (DISs) from rough set theory [36], optimizing CRP via three key points. First, it replaces manual threshold settings with multidimensional analysis to objectively quantify consensus and conflict levels. Second, it employs attribute-level diagnostics to precisely identify evaluation criteria causing divergence. Lastly, it establishes a "gradient consensus" mechanism that balances decision efficiency with the preservation of minority opinions' innovative potential. This design addresses limitations in traditional methods, offering a more flexible solution for complex decision scenarios.

In real-world decision contexts, consensus-driven approaches may inadvertently suppress dissenting voices, as evidenced by historical cases where breakthrough innovations predominantly originated from nonconforming perspectives (e.g., the Copernican revolution). Minority opinions in LSGDM frequently play a crucial role, particularly in guiding the group toward innovative thinking and exploration [37]. The risks associated with safeguarding minority opinions mainly manifest in SNs and group dynamics. Complex mechanisms are required in LSGDM to identify and evaluate minority opinions.

Ren et al. [38] suggested using an 18% threshold to define minority groups in microgrid planning. Xu et al. [39] proposed a subgroup-quantification mechanism that systematically evaluates both the cardinality and influence of the weight of minority-advocating subgroups within CRPs, particularly addressing the challenge of preserving valuable dissent while maintaining decision efficiency. They adjusted subgroup weights based on the value of minority opinions, achieving more reasonable solutions. While these studies provide valuable insights, they have yet to address the subjectivity involved in identifying minority opinions.

This study proposes a dynamic optimization framework for minority opinions based on group trust. By employing social influence analysis and logical verification to precisely identify high-value dissenting opinions, and utilizing prospect–regret theory to intelligently balance opinion refinement with innovation preservation, the framework enhances decision-making contributions while ensuring consensus efficiencies, thereby overcoming the limitations of traditional approaches.

### *1.4. Case Study Context in UAVs*

UAVs, as a rapidly evolving technology, have been widely adopted across diverse domains such as logistics, surveillance, and environmental monitoring. The selection of UAVs involves intricate decision-making processes due to the multiplicity of attributes (e.g., flight endurance, payload capacity, and cost), the heterogeneity of stakeholders (e.g., manufacturers, operators, and regulators), and the introduction of massive opinions and dynamic trust relationships via online reviews and social media platforms [4]. Addressing these challenges necessitates robust LSGDM frameworks capable of integrating multi-source data, reconciling conflicting preferences, and ensuring consensus.

To operationalize sentiment-aware decision modeling, textual reviews from platforms like Amazon and specialized UAV forums are collected and preprocessed. Sentiment scores $s \in [-1, 1]$ are computed using a lexicon-based approach combined with bidirectional encoder representations from transformers fine-tuned for product reviews. This study employs UAV online reviews as a real-world case study to validate the proposed model, demonstrating its applicability in practical decision scenarios and bridging the gap between theoretical LSGDM research and real-world implementation.

### *1.5. Summary of Our Method*

Building upon the limitations highlighted in previous LSGDM approaches, this paper aims to address several key objectives:

(1) In SN-LSGDM, DMs may be influenced by their social neighbors or experience hesitation during the evaluation process. Therefore, examining their behavioral patterns within SNs is crucial. While various trust propagation operators have been proposed to enhance the decision process, existing methods for handling missing values often rely on simple statistical estimations, which introduce bias and lack interpretability. Hence, developing a propagation algorithm that not only improves the decision accuracy, but also enhances the interpretability is essential for completing trust matrices.

(2) Most existing clustering methods [27] primarily focus on a single influencing factor, which often leads to suboptimal or even contradictory clustering results. To overcome this limitation, it is necessary to design a more robust clustering method that integrates two key influencing factors, ensuring greater accuracy and consistency in clustering outcomes.

(3) The consensus measurement plays a pivotal role in CRP. Current methods typically determine consensus levels by setting predefined thresholds, which are often sub-

jective. Therefore, developing a more objective and formula-driven measurement approach is essential for improving fairness and reliability in consensus assessment.

(4) Existing methods for managing minority opinions generally rely on analyzing subjective discussions to assess their significance. However, given the inherent uncertainty and risk in real-world decision-making, a trust-based detection and weight-management approach can provide a more effective solution for evaluating and integrating minority opinions in the decision process.

As shown in Figure 1, our proposed framework outlines the systematic process as follows. The entire process begins with data intuitionistic fuzzification, where data are transformed into IF matrices, which are the evaluation matrices of DMs. Secondly, an enhanced K-means++ clustering method is adopted. The formula for hybrid distances is defined, and then K-means++ clustering is carried out to form subgroups. For each subgroup, subgroup-evaluation matrices will be created. The process also involves SNA. The adjacency matrices are processed through the Hamacher aggregation operator with a decay factor to derive trust matrices. The degree centrality index is calculated, providing a basis for weight calculation and trust propagation. Then, by using the IF lower approximation and upper approximation, decision index sets $T1$, $T2$, and $T3$ are formed. A consensus measure based on DISs is calculated. The decision-making step checks for minority opinions. If a large-scale conflict is detected, an identification mechanism is triggered. Depending on the result, either $M1$ or $M2$ is applied. If there is no large-scale conflict, a small-scale conflict is checked. For each alternative, scores are obtained from $T3$, and the final ranking is determined through the selection process. This systematic flowchart integrates multiple techniques and can effectively handle LSGDM problems.

The key contributions and innovations of this paper are summarized as follows:

(1) The completion of incomplete trust matrices: combining the Hamacher aggregation operator with a decay factor to handle missing values in incomplete trust matrices.

(2) Enhanced K-means++ clustering: integrating both cardinal and ordinal distance metrics to develop a personalized distance formula for the classical K-means++ algorithm. An advanced weighting method is proposed, considering the number of DMs and trust levels within the community.

(3) The objective consensus measurement: introducing a DIS for conflict detection, offering a high level of objectivity and fairness in determining whether consensus has been reached.

(4) The management of minority opinions: developing an improved method for identifying and adjusting minority opinions, leveraging group trusts to assess their importance and adjusting their weights accordingly.

The remainder of this paper is structured as follows: Section 2 introduces key concepts and methods discussed in this research. Section 3 describes the data-processing workflow and the construction of SNs. Section 4 details the clustering and weighting methods. Section 5 focuses on the consensus-reaching method based on rough decision indicators and minority opinions. In Section 6, a case study of UAV online reviews for selection is used to validate the model's feasibility, with sensitivity analysis and comparison experiments. Finally, the paper concludes with a summary and outlook for future research.

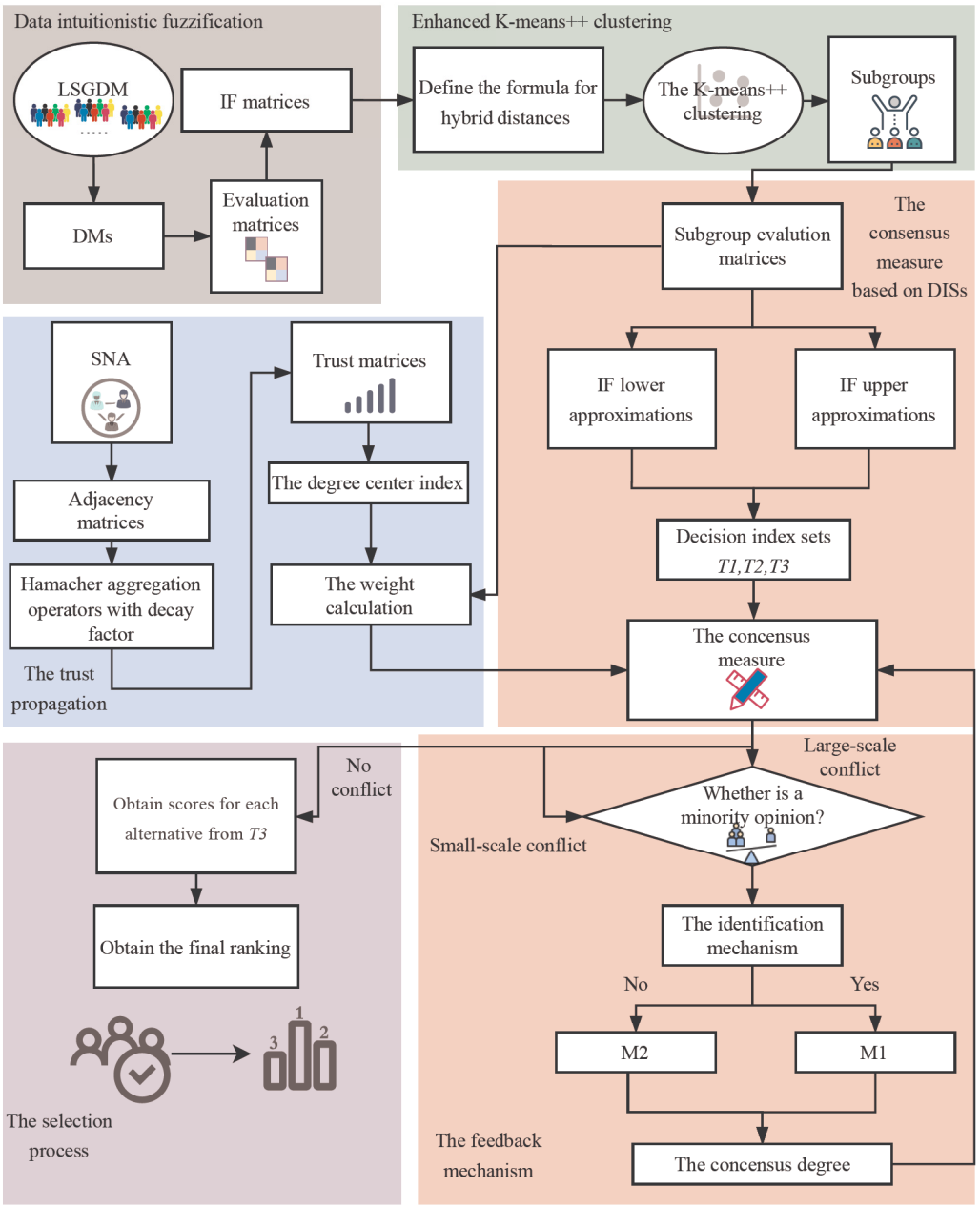



**Figure 1.** LSGDM flowchat.

## 2. Preliminaries

This section first introduces the framework of the LSGDM model presented in this study. It then reviews intuitionistic fuzzy numbers (IFNs), the trust relationships under SNA, and their propagation operators. Finally, it provides a detailed introduction to minority opinion handling (MOH) and prospect–regret theory. These foundational concepts will aid in understanding the proposed method.

### *2.1. A Dynamic Consensus-Driven Framework for LSGDM*

As application scenarios evolve, the decision-making process integrates various evaluation methods to better address increasing complexity and uncertainty. This section focuses on some of the basic concepts and main processes associated with LSGDM to gain a deeper understanding of LSGDM. More detailed and relevant symbols are introduced as shown in Table 1.

**Table 1.** Notations and related descriptions.

| Mathematical Symbols | Meaning |
|---|---|
| $E = \{e_1, e_2, \cdots, e_d\}$ | The set of experts |
| $Q_k = (q_{ij}^k)_{n\times m}$ | The decision-making matrix of a DM $e_k$ |
| $A = \{a_1, a_2, \cdots, a_m\}$ | The set of attributes |
| $O = \{o_1, o_2, \cdots, o_m\}$ | Attributes weights |
| $X = \{x_1, x_2, \cdots, x_n\}$ | The set of alternatives |
| $\mathrm{SM} = (sm_{kh})_{d\times d}$ | The trust matrix |
| $\mathrm{FS} = (fs_{kh})_{d\times d}$ | The fuzzy number trust degree matrix |
| $\mathrm{TS}_h$ | The credibility measure of a subgroup |
| $\mathrm{SC}_{ij}^k$ | The score value related to a DM $e_k$ |
| $\alpha = (\mu_\alpha, \nu_\alpha)$ | The representation form of an IFN |
| $C = \{C_1, C_2, \cdots, C_b\}$ | The set of subgroups |
| $fp$ | The index related to positive certainty |
| $rp$ | The index related to negative suspicion |
| $C_i = (c_{ij}^k)_{n\times m}$ | The evaluation matrix after subgroup aggregations |

The methodological framework for LSGDM problem-solving is structured as follows:

**Step 1: Data processing.** In real-world decision contexts, the ontological fuzziness (inherent unobservability of essence) and epistemological uncertainty (cognitive limitations in subjective judgment) jointly constitute dual cognitive barriers. Intuitionistic fuzzy set (IFS) theory addresses such bimodal uncertainties via constructing a triple-measure space $\mu, \nu, \pi$ satisfying $\mu + \nu + \pi = 1$. For sentiment scores $s \in [-1, 1]$ derived from textual analysis, the bijective mapping rule is formalized as:

(1) The meaning of sentiment scores: let the raw sentiment score $s \in [-1, 1]$, where $s = 1$ indicates extreme the positivity and $s = -1$ indicates extreme the negativity.
(2) The decomposition of bipolar affective measures: positive certainty: $fp = \frac{s+1}{2}$; negative skepticism: $rp = 1 - fp$.
(3) The quantification of cognitive uncertainty: certainty index (confidence in affective judgment): $d = |s|$.
(4) The synthesis of IFNs: the final IFN triplet is constructed via certainty-modulated fusion: $\mu = fp \times d; \nu = rp \times d$.

**Step 2: Trust propagation.** Within the context of socio-structural evolution driving enhanced interpersonal connectivity, the trust diffusion mechanism employs Hamacher aggregation operators to model trust relationships, thereby effectively establishing trust links between initially disconnected nodes.

**Step 3: The clustering process.** To minimize redundant calculations and reduce the complexity of LSGDM problems, DMs with similar viewpoints are grouped into distinct subgroups. Consider an large-scale group partitioned into $b$ disjoint subgroups, denoted as $C = \{C_1, C_2, \cdots, C_b\}$. The subgroup decision matrices are subsequently derived via the weighted average (WA) operator [40]. Specifically, the decision matrix for the subgroup $C_h$, represented as $V_h = \left(v_{ij}{}^h\right)_{n\times m}$, is obtained via the WA calculation.

**Step 4: The CRP.** The consensus-reaching mechanism within group-based structures should be grounded in intra-group collaboration, subsequently achieving hierarchical integration via inter-group negotiation and adjustment mechanisms, thereby ensuring broad acceptance of decision outcomes at the macro-group level.

**Step 5: Obtaining final ranking.** Once the group reaches the predefined consensus level, the final group decision matrix $G = \left(g_{ij}\right)_{n\times m}$ is obtained. To rank all alternatives, the score value for each alternative is calculated based on the method outlined in the following Equation (1):

$$SC_{ij}{}^{k} = \mu_{ij}{}^{k} - v_{ij}{}^{k}. \tag{1}$$

*2.2. IFNs*

This section primarily introduces the basic definition, operational rules, and comparison rules of IFNs.

2.2.1. The Basic Concept of IFNs

IFNs serve as an important extension for classic fuzzy numbers, designed to more accurately represent uncertainty and fuzziness. The key feature of IFNs lies in the introduction of three fundamental concepts: membership degrees, non-membership degrees, and hesitation degrees. These elements collectively provide a more comprehensive representation of the degree of fuzziness and uncertainty in fuzzy sets.

**Definition 1.** *Let $X$ be a non-empty classical set, then an IFS on $X$ can be defined as Equation (2) [41]:*

$$\widetilde{A} = \{\langle x, \mu_{\widetilde{A}}(x), \nu_{\widetilde{A}}(x)\rangle | x \in X\}, \tag{2}$$

*where $\mu_{\widetilde{A}}(x)$ and $\nu_{\widetilde{A}}(x)$ represent the membership degree and non-membership degree of the element $x$ belonging to the IFS $\widetilde{A}$, respectively, as shown in the following Equations (3) and (4).*

$$\mu_{\widetilde{A}} : X \to [0,1], x \in X \to \mu_{\widetilde{A}}(x) \in [0,1], \tag{3}$$

$$\nu_{\widetilde{A}} : X \to [0,1], x \in X \to \nu_{\widetilde{A}}(x) \in [0,1], \tag{4}$$

*where $0 \leq \mu_{\widetilde{A}}(x) + \nu_{\widetilde{A}}(x) \leq 1, x \in X$.*

For convenience, $\alpha = (\mu_\alpha, \nu_\alpha)$ is directly referred to as an IFN, satisfying the condition $0 \leq \mu_\alpha(x) \leq 1$, $0 \leq \nu_\alpha(x) \leq 1$, $0 \leq \mu_\alpha(x) + \nu_\alpha(x) \leq 1$.

2.2.2. IF Operations

**Definition 2.** *Let $\widetilde{A} = \{< x, \mu_{\widetilde{A}}(x), \nu_{\widetilde{A}}(x) > | x \in X\}$ and $\widetilde{B} = \{< x, \mu_{\widetilde{B}}(x), \nu_{\widetilde{B}}(x) > | x \in X\}$ be two IFSs on $X$ [41]. Then:*

*(1) The sum of IFSs, as shown in Equation (5):*

$$\widetilde{A} + \widetilde{B} = \{< x, \mu_{\widetilde{A}}(x) + \mu_{\widetilde{B}}(x) - \mu_{\widetilde{A}}(x)\mu_{\widetilde{B}}(x), \nu_{\widetilde{A}}(x)\nu_{\widetilde{A}}(x) > | x \in X\}. \tag{5}$$

*(2) The product of IFSs, as shown in Equation (6):*

$$\widetilde{A} \times \widetilde{B} = \{< x, \mu_{\widetilde{A}}(x)\mu_{\widetilde{B}}(x), \nu_{\widetilde{A}}(x) + \nu_{\widetilde{B}}(x) - \mu_{\widetilde{A}}(x)\mu_{\widetilde{B}}(x) > | x \in X\}. \tag{6}$$

*(3) The multiplication of an IFS with a scalar, as shown in Equation (7):*

$$\lambda\widetilde{A} = \{< x, 1 - (1 - \mu_{\widetilde{A}}(x))^{\lambda}, (\nu_{\widetilde{A}})^{\lambda} > | x \in X\}. \tag{7}$$

2.2.3. IF Comparison Rules

**Definition 3.** *For an IFS $\alpha = (\mu_\alpha, v_\alpha)$, its score function is defined as the following Equation (8):*

$$SC(\alpha) - \mu_\alpha - v_\alpha, \tag{8}$$

*its exact function is defined as Equation (9):*

$$H(\alpha) = \mu_\alpha + v_\alpha. \tag{9}$$

*Based on Equations (8) and (9), some schemes for comparing two IFNs* $\alpha_i = (\mu_{\alpha_i}, v_{\alpha_i}), \alpha_j = (\mu_{\alpha_j}, v_{\alpha_j})$ *are as Equations (10) and (11).*

**Rule 1.**

$$\text{If } \mathrm{SC}(\alpha_\mathrm{i}) < \mathrm{SC}(\alpha_\mathrm{j}), \text{ then } \alpha_\mathrm{i} < \alpha_\mathrm{j}; \tag{10}$$

$$\begin{aligned} &\text{If } \mathrm{SC}(\alpha_\mathrm{i}) = \mathrm{SC}(\alpha_\mathrm{j}) \text{ then:} \\ &\quad \text{if } \mathrm{H}(\alpha_\mathrm{i}) < \mathrm{H}(\alpha_\mathrm{j}) \text{ then } \alpha_\mathrm{i} < \alpha_\mathrm{j}; \\ &\quad \text{if } \mathrm{H}(\alpha_\mathrm{i}) = \mathrm{H}(\alpha_\mathrm{j}) \text{ then } \alpha_\mathrm{i} = \alpha_\mathrm{j}. \end{aligned} \tag{11}$$

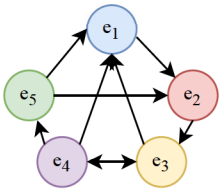
### *2.3. SNA*

This subsection primarily introduces three different representations of trust networks and the Hamacher aggregation operator.

#### 2.3.1. The Basic Concept of SNs

SNA constitutes a computational paradigm for modeling relational structures among social actors. This methodology constructs group relational networks to uncover three-tier structural characteristics [42]: (1) macroscopic social topology, (2) mesoscopic node centrality metrics, and (3) microscopic individual behavioral dynamics. The standardized SNA framework comprises three fundamental elements: first, node sets representing heterogeneous social agents (e.g., individuals, organizations, nation-states); second, edge sets defining multiplex relational ties (including affective bonds, trust relationships, collaborative linkages); third, attribute vectors characterizing multidimensional nodal features (e.g., demographic variables, social capital indices, preference parameters). These elements can be systematically represented via a multilayer network framework as demonstrated in Table 2.

**Table 2.** SN representational taxonomies.

| Graph | Algebraic | Sociometric |
|---|---|---|
|   | $e_1\mathrm{Re}e_2, e_2\mathrm{Re}e_3, e_3\mathrm{Re}e_4,$ $e_3\mathrm{Re}e_1, e_4\mathrm{Re}e_5, e_5\mathrm{Re}e_1,$ $e_5\mathrm{Re}e_2, e_4\mathrm{Re}e_1, e_4\mathrm{Re}e_3$ | $\begin{bmatrix} - & 1 & 0 & 0 & 0 \\ 0 & - & 1 & 0 & 0 \\ 1 & 0 & - & 1 & 0 \\ 1 & 0 & 1 & - & 1 \\ 1 & 1 & 0 & 0 & - \end{bmatrix}$ |

(1) **The graph representation:** A graph is used to model the trust network among members, where nodes represent individuals, and directed edges denote trust relationships. Specifically, an edge $e_1 \to e_3$ indicates a direct trust relationship from $e_1$ to $e_2$.

(2) **The algebraic representation:** This method differentiates various types of relationships and demonstrates how these relationships can be combined for analysis.

(3) **The sociometric representation:** Trust relationships between members are expressed in a matrix $SM = (sm_{kh})_{d\times d}$, where each element indicates the presence or absence of a direct trust relationship. Specifically, if $sm_{kh} = 1$, it represents a direct trust relationship from $e_k$ to $e_h$. Conversely, if $sm_{kh} = 0$, it signifies the lack of such a relationship.

The aforementioned representations are limited to explicit trust relationships, which imply either full trust or complete distrust. However, in real-world scenarios, trust relationships among DMs are often ambiguous and cannot be described with absolute certainty. To address this, Zhang et al. [43] devised a granular computing-based social trust-quantification paradigm, enabling DMs to explicitly quantify interpersonal trust

degrees via unit interval [0,1]-valued mappings. In this framework, the trust relationships within a social trust network are described using the fuzzy sociometric representation.

**Definition 4.** *Let the fuzzy social trust matrix be* $FS = (fs_{kh})_{d\times d}$ *over an expert set* $E$ *with* $n = |E|$. *Its membership function is denoted by* $\eta_t : E \times E \to [0,1]$, *where* $\eta_t(e_k, e_h) = fs_{kh}$, *and* $fs_{kh}$ *indicates the trust level assigned by* $e_k$ *to* $e_h$ *[42].*

**Example 1.** *The trust levels among five DMs, are represented by an incomplete fuzzy sociometric matrix* $FS = (fs_{kh})_{d\times d}$:

$$FS = \begin{pmatrix} - & 0.90 & 0 & 0 & 0 \\ 0 & - & 0.75 & 0 & 0 \\ 0.85 & 0 & - & 0.65 & 0 \\ 0.88 & 0 & 0.60 & - & 0.83 \\ 0.70 & 0.55 & 0 & 0 & - \end{pmatrix}.$$

In social trust networks, the direct observation of trust values between specific DMs often encounters data incompleteness, resulting in a partially defined fuzzy adjacency matrix. As illustrated in Figure 2 depicting a triadic network structure comprising three DMs, the direct trust relationship between $e_1$ and $e_3$ is unobserved. This structural gap necessitates applying trust transitivity principles, where the mediated path $(e_1 \to e_2 \to e_3)$ facilitates the probabilistic inference of indirect trust propagations.

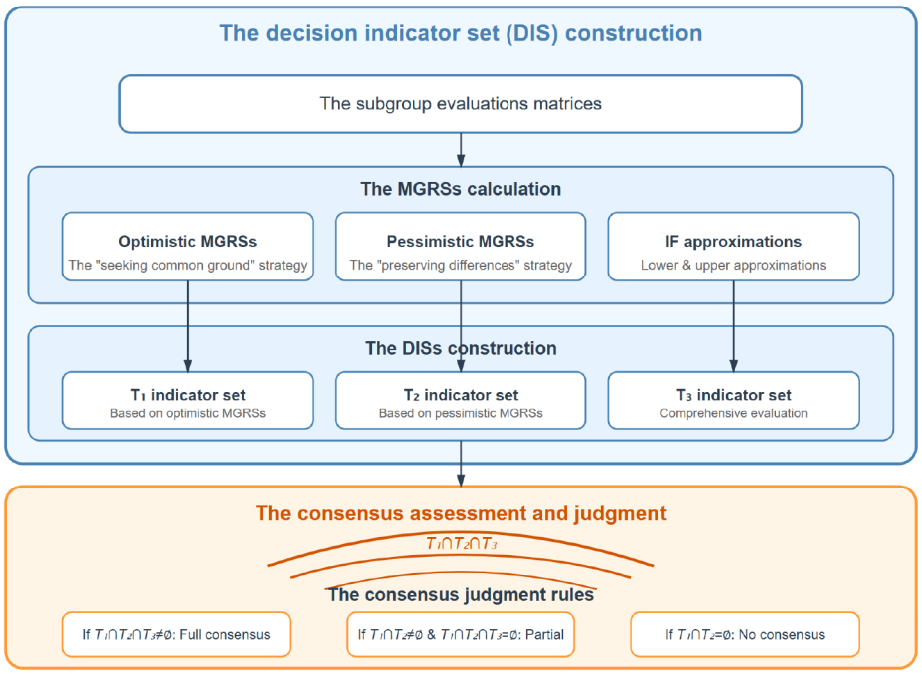



**Figure 2.** The DIS consensus-measurement process.

### 2.3.2. Hamacher Aggregation Operator

Before explaining our trust propagation method, it is important to introduce key concepts related to triangular norms. define the generalized intersection and union of fuzzy sets. A more flexible approach was introduced by generalizing t-norm operations, leading to the development of Hamacher operations, specifically the Hamacher product $\otimes$,

which are broader formulations of t-norms [44,45]. Thus, these operations are referred to as Hamacher t-norms. The mathematical definition of this operation is as Equation (12):

$$T(a_1, a_2) = a_1 \otimes a_2 = \frac{a_1 a_2}{\rho + (1-\rho)(a_1 + a_2 - a_1 a_2)}, \rho > 1, \tag{12}$$

where $\rho$ is the weighted product and adjustment factor, which controls the interaction between the two values and plays a key role in adjusting their impact on the final aggregated value.

In order to incorporate $n$ independent variables, the equation is extended via the use of Equation (13):

$$T(a_1, a_2, \cdots, a_n) = \frac{\prod_{i=1}^{n} a_i}{\rho + (1-\rho)\sum_{i=1}^{n} a_i - \prod_{i=1}^{n} a_i}. \tag{13}$$

**Remark 1.** *Pilot experiments with UAV reveal that $\alpha = 0.8$ strikes an optimal balance between retaining the reasonable influence of historical information and adapting to current changes. It gradually reduces the impact of historical information on the current iteration's results, ensuring the stability and convergence of the trust propagation process. Specifically, when compared with other potential values (e.g., $\alpha$ = 0.5 or $\alpha$ = 0.9), $\alpha = 0.8$ minimizes the convergence error in trust-value calculations.*

**Remark 2.** *To ensure that all missing values in the final trust matrix are effectively filled, any positions that remain labeled as "none" after trust propagation are assigned a default value of 0.5. This choice reflects a neutral, compromise-level trust score, which-absent more precise information-provides a reasonably fair estimate for subsequent analyses. As a result, the trust matrix becomes fully populated and is ready for further in-depth analysis and presentation.*

**Remark 3.** *During the propagation process, for each missing trust value (i.e., none), the algorithm propagates trust from neighboring nodes (other evaluators) based on the weight values in the weight matrix, in order to estimate the missing values and gradually refine the trust distribution.*

### *2.4. Theoretical Synthesis of Prospect and Regret Mechanisms*

The prospect–regret dual decision framework establishes a robust analytical framework for modeling cognitive biases in behavioral decision-making. This innovative paradigm synthesizes two cornerstone theories: prospect theory that accounts for reference-dependent preferences under risk, and regret theory that formalizes anticipated emotional responses to counterfactual outcomes. Through this theoretical integration, the framework particularly elucidates two pervasive behavioral anomalies in financial decision contexts: (1) herd mentality driven by social reference dependence, and (2) systematic distortion of probability weighting in risk-appraisal processes [46,47].

**Definition 5.** *The axiomatic foundation of this framework is formalized via a value function with distinct mathematical representations for gain and loss domains [48], as shown in Equation (14):*

$$v(\Delta x_{ij}) = \begin{cases} (\Delta x_{ij})^{\alpha}, & \Delta x_{ij} \geq 0 \\ -\lambda(-\Delta x_{ij})^{\beta}, & \Delta x_{ij} < 0 \end{cases}. \tag{14}$$

**Definition 6.** *The regret–rejoice function. is mathematically characterized by [49], as shown in Equation (15):*

$$R(\Delta v) = 1 - e^{-\delta \Delta v}, \quad \delta > 0, \tag{15}$$

*where* $\Delta v = v(x) - v(y)$ *represents the utility differential between alternatives* $X$ *and* $Y$*. The function exhibits:*

- *Monotonicity:* $R'(\Delta v) \geq 0$.
- *Diminishing sensitivity:* $R''(\Delta v) < 0$.
- *Sign-dependent interpretation:* $R(\Delta v) \geq 0$ *signifies rejoicing, otherwise regret.*

**Definition 7.** *The perceived utility synthesizes intrinsic value and counterfactual emotion, as shown in Equation (16):*

$$U(X) = v(x) + R(\Delta v), \tag{16}$$

*where* $v : \mathcal{X} \to \mathbb{R}$ *maps alternatives to cardinal utilities, and* $R(\cdot)$ *modulates emotional responses.*

Table 3 presents the preset values of each constant and their corresponding behavioral interpretations.

**Table 3.** Key parameters.

| Symbols | Keys | Behavioral Interpretations |
|---|---|---|
| $\alpha$ | 0.88 | Risk sensitivity in gains |
| $\beta$ | 0.92 | Loss aversion magnitude |
| $\lambda$ | 2.25 | Loss amplification factor |
| $\delta$ | 0.2 | Regret aversion intensity |

## 3. Clustering and Calculating Weights

This section focuses on the detailed implementation process of the enhanced K-means++ clustering algorithm and provides a thorough explanation of how the weights for each evaluator and group are calculated based on trust values.

### *3.1. Enhanced K-means++ Clustering*

The enhanced K-means++ clustering proposed in this paper is an improvement upon the traditional K-means++ method, which utilizes the elbow method to determine the optimal number of clusters, $k$. Specifically, the optimal $k$ is selected by plotting the relationship between the sum of squared errors (SSE) and the number of clusters, choosing the "elbow" point where the rate of decrease in SSE begins to slow down. Additionally, the algorithm optimizes the selection of initial centroids and distance calculations. In the distance computation, it considers not only the DMs' evaluation of the alternatives but also takes into account their preferences for ranking these alternatives, while incorporating SNs. By further clustering the DMs in multiple dimensions, the algorithm enhances the accuracy and effectiveness of the decision process.

#### 3.1.1. A I Iybrid Distance Metric Integrating Cardinal and Ordinal Distances

Conventional LSGDM divergence metrics predominantly operate on the value space. This section develops a novel multidimensional divergence metric that simultaneously accounts for both numerical evaluation discrepancies and preferential ranking interdependencies among DMs. Operationally, we integrate two complementary measurement paradigms: continuous-space metrics (Euclidean/Manhattan norms) to quantify value-based discrepancies in evaluative judgments, coupled with discrete metric analysis (Hamming distances) to assess ordinal mismatches in alternative preference orderings.

**Definition 8.** *Let* $Q_{k1} = (q_{ij}{}^{k1})_{n\times m}$ *and* $Q_{k2} = (q_{ij}{}^{k2})_{n\times m}$ *be evaluation matrices of DMs* $e_{k1}, e_{k2}$*, where* $e_{k1}, e_{k2} \in E$*. Using the Euclidean distance, the distance between them is calculated as Equation (17) [30]:*

$$d_1(Q_{k1}, Q_{k2}) = \sum_{i=1}^{n}\sum_{j=1}^{m}\sqrt{(q_{ij}{}^{k1} - q_{ij}{}^{k2})^2}, \forall j \in M, i \in N, k1, k2 \in G. \tag{17}$$

**Definition 9.** *Let* $Q_{k1} = (q_{ij}{}^{k1})_{n\times m}$ *and* $Q_{k2} = (q_{ij}{}^{k2})_{n\times m}$ *be evaluation matrices of DMs* $e_{k1}, e_{k2}$*, where* $e_{k1}, e_{k2} \in E$*. Using the Manhattan distance, the distance between them is calculated as in Equation (18) [27]:*

$$d_2(Q_{k1}, Q_{k2}) = \sum_{i=1}^{n}\sum_{j=1}^{m}|q_{ij}{}^{k1} - q_{ij}{}^{k2}|, \forall j \in M, i \in N, k1, k2 \in G. \tag{18}$$

Prior to computing the Hamming distance, a necessary precondition is to algorithmically construct a total order of alternatives for each DM via systematic processing of the evaluation matrix, with explicit tie-breaking rules for handling equivalent preference scores.

**Definition 10.** *Let* $Q_k = (q_{ij}{}^{k})_{n\times m}$ *be an evaluation matrix of any DM* $e_k$ *and* $w_j$ *a weight of the j-th attribute, then the DM's evaluation value* $VA_k$ *for each option is as in Equation (19) [30]:*

$$VQ_k = \sum_{j=1}^{m} w_j q_{ij}{}^{k}, \forall j \in M, i \in N, \tag{19}$$

*in this context,* $VQ_k$ *represents the ranking assessment values given by the k-th DM for each option. By ordering the values in* $VQ_k$ *from highest to lowest, we can derive the ranking of alternatives for the k-th DM, denoted by* $VA_k$*.*

Upon establishing the complete ordinal preference structures for all DMs via total ordering transformation, an enhanced Hamming distance metric can be applied to quantify inter-DM ordinal space dispersion. This measure operationalizes cognitive divergence by identifying the number of discrepancy loci between two ranking vectors, formally defined as follows.

**Definition 11.** *Let* $VA_{k1}$ *and* $VA_{k2}$ *denote two alternatives ranking vectors from distinct DMs k*1*, k*2*, The Hamming distance between these ordinal rankings is defined as Equation (20) [30]:*

$$d_3(VA_{k_1}, VA_{k_2}) = \sum_{h=1}^{n}\mathbb{I}(VA_{k_1}(h) \neq VA_{k_2}(h)), \tag{20}$$

*where:*

- $h \in \{1, 2, \cdots, n\}$ *indexes the ordinal position in the ranking;*
- *The indicator function* $\mathbb{I}(\cdot)$ *is defined as:*

$$\mathbb{I}\left(v_{k_1}^{(h)} \neq v_{k_2}^{(h)}\right) = \begin{cases} 1 & if\, v_{k_1}^{(h)} \neq v_{k_2}^{(h)} \\ 0 & otherwise \end{cases}.$$

The distance formulas discussed above focus on capturing the connection between DMs' evaluation values and their resulting rankings of alternatives. Building on this foundation, the subsequent analysis integrates both dimensions—numerical assessments and final rankings—into a unified framework. This approach ensures a comprehensive view of how DMs' evaluations translate into preference orders.

**Definition 12.** *Let $Q_{k_1} = (q_{ij}^{k_1})_{n\times m}$ and $Q_{k_2} = (q_{ij}^{k_2})_{n\times m}$ denote evaluation matrices DMs $k_1$ and $k_2$, respectively, with corresponding alternative ranking vectors $VA_{k_1}$ and $VA_{k_2}$. The composite distance metric integrating both evaluation and ranking similarities is formulated as Equation (21):*

$$d_4(Q_{k1}, Q_{k2}) = \alpha\left(\frac{1}{n}\sum_{h=1}^{n}\mathbb{I}(VA_{k_1}(h) \neq VA_{k_2}(h))\right) + (1-\alpha)\left[\beta\left(\sum_{i,j}(q_{ij}^{k1} - q_{ij}^{k2})^2\right)^{1/2} + (1-\beta)\sum_{i,j}|q_{ij}^{k1} - q_{ij}^{k2}|\right], \quad \text{for all } i \in N,\ j \in M,\ k1, k2 \in G. \tag{21}$$

*where $\alpha, \beta \in [0, 1]$ are weighting parameters (default: $\alpha = \beta = 0.5$).*

The parameters $\alpha$ and $\beta$ in the hybrid distance metric were set to default values of 0.5 based on the principle of balanced contribution between ranking and evaluation similarities.

Specifically: $\alpha = 0.5$ ensures equal emphasis on ranking divergence (via normalized Hamming distance) and numerical evaluation divergence (via Euclidean/Manhattan distances). This design acknowledges that both ordinal preferences and cardinal evaluations are critical in capturing DMs' behavioral duality. $\beta = 0.5$ assigns equal weight to Euclidean and Manhattan distances, as both metrics are widely recognized for their complementary strengths in quantifying numerical discrepancies (Euclidean for geometric dissimilarity, Manhattan for grid-like deviations).

**Remark 4.** *The distance formula proposed in Definition 12 combines the Euclidean distance with the Manhattan distance to jointly measure the similarity of the evaluation information between DMs, while also integrating the Hamming distance to evaluate the similarity between alternative ranking outcomes. To effectively balance the contributions of these three distances, the Hamming distance is normalized to ensure it is on the same scale as the Euclidean and Manhattan distances, creating a unified metric system.*

**Remark 5.** *Additionally, two parameters are introduced to further adjust the relative impact of these three distances in the final result. These parameters not only allow for flexible weight adjustment of each distance but also optimize the distance calculation based on specific circumstances, improving the model's applicability and accuracy.*

#### 3.1.2. K-means++ Clustering

The original K-means algorithm randomly selects $k$ data points from the dataset as initial cluster centers. However, the selection of these initial centroids significantly impacts both the final clustering results and computational efficiency. A purely random initialization may lead to slow convergence. The K-means++ algorithm addresses this limitation by optimizing the initialization process for selecting initial centroids.

The core principle of K-means++ lies in maximizing the distances between initial cluster centers to mitigate suboptimal convergence. Its initialization procedure operates as follows: (1) first cluster center: randomly select one data point from the dataset; (2) subsequent cluster centers: for each remaining data point, calculate its shortest distance to any existing cluster center. New centers are then selected based on a distance-dependent probability distribution, where the greater the distance of a data point from existing centers, the higher its probability of being chosen as the next cluster center. This process repeats until $k$ initial centroids are determined. Finally, the standard K-means algorithm iteratively refines these initial centers to obtain the final clustering solution.

**Definition 13.** *Let $C_i$ be the i-th cluster, q be the sample point in $C_i$, and $m_i$ be the particle of $C_i$ (the mean of all samples in $C_i$), which represents the quality of the clustering effect, as shown in Equation (22).*

$$SSE = \sum_{i=1}^{k} \sum_{q \in C_i} |q - m_i|^2. \tag{22}$$

Additionally, the selection of $k$ value adopts the elbow method. In the process of clustering, with the increase of the number of clusters, the sample division will become more refined, the degree of aggregation of each category is higher, then the SSE will gradually become smaller, the sum of squares of error is the sum of squares of the distance between the center of gravity of this class and the position of its internal members. SSE is the core index of the elbow method. When $k$ is less than the real cluster number, SSE will decrease greatly because the increase of $k$ will increase the degree of aggregation of each cluster. When $k$ reaches the real number of clusters, the return on the degree of aggregation obtained by increasing $k$ will quickly become smaller, thus the decline of SSE will sharply decrease, and then it will flatten out with the continuous increase of $k$ value. That is to say, the relationship between SSE and $k$ is the shape of an elbow, and the corresponding $k$ value of this elbow is the real cluster number of data.

The complete implementation workflow of the proposed enhanced K-means++ clustering algorithm is formally presented in Algorithm 1.

**Algorithm 1** Enhanced K-means++ clustering with hybrid features

**Input:** The DM evaluation matrix set $\{Q_k\}_{k=1}^{d}$, the attribute weight $W$.
**Output:** The subgroup partition result $\{C_h\}_{h=1}^{b}$.
**Step 1: Data preprocessing**
Compute the alternative sorting vector $\{VA_k\}$ for all DMs using Equation (19);
Construct the mixed feature space: $\mathbb{F} = \mathbb{R}^{n \times m} \times \mathbb{N}^n$;
**Step 2: Determine the number of clusters**
Initialize $k_{\max} = 10$, the SSE list;
**for** $k = 2$ **to** $k_{\max}$ **do**
    Perform standard K-means++ clustering;
    Compute within-cluster squared error using Equation (22);
**end for**
Determine the optimal $k$ value using the elbow method;
**Step 3: Mixed distance clustering**
Initialize cluster centers (density peak selection);
**while** not converged **do**
    **for** each DM $e_k$ **do**
        Compute the mixed distance to each cluster center using Equation (21);
        Assign to the nearest cluster;
    **end for**
    Update cluster centers (weighted IF averaging);
**end while**

### *3.2. Weight Determination Framework*

The traditional majority principle typically calculates weights based on the number of DMs within a subgroup, and this method has been widely applied. However, it overlooks other important characteristics within the subgroup, such as the diversity of opinions, the experience of DMs, and their influence on the decision process. Therefore, relying solely on the number of individuals to determine the weights has its limitations. As a result, there is an urgent need to develop a more comprehensive weight determination method that considers these additional factors to improve the accuracy and fairness of the weight assignment process [30].

**Definition 14.** *Let the sociomatrix of a trust network be* $FS = (fs_{ij})_{d\times d}$. *where* $fs_{ij}$ *quantifies the directed trust intensity from* $e_i$ *to* $e_j$. *The normalized degree centrality measure* $CI_k$ *for the DM* $e_k$ *is calculated as in Equation (23) [44]:*

$$CI_k = \frac{1}{d-1}\sum_{i=1,i\neq k}^{d} fs_{ik}. \tag{23}$$

Within SNA, the degree centrality metric serves as a quantifiable proxy for evaluating nodal influence, wherein elevated centrality values correlate with enhanced structural prominence of decision agents. This property establishes degree centrality as a principal determinant in nodal weighting schemes, given its capacity to systematically prioritize stakeholders exhibiting maximal social capital accumulation via trust linkage density.

Then, the weight $w_k$ of the DM $e_k$ can be defined as Equation (24):

$$w_k = \frac{CI_k}{\sum_{i=1}^{d} CI_k}. \tag{24}$$

**Definition 15.** *The credibility* $TS_h$ *of the subgroup* $C_h$ *is computed as the following Equation (25) [42]:*

$$TS_h = \frac{1}{d-\#C_h}\cdot\frac{1}{\#C_h}\sum_{i=1,i\neq k}^{d} fs_{ik}, \tag{25}$$

*where* $\#C_h$ *is the number of DMs within* $C_h$, $s_{ik}$ *represents the trust degree assigned by* $e_i$ *to* $e_k$.

**Definition 16.** *By calculating* $TS_h$ *that can be applied to measure the credibility of* $C_h$, *and the DMs in* $C_h$ *that can also be applied to reflect the significance of the subgroup, we define the weights of the subgroup* $C_h$ *by integrating these two indicators as in Equation (26):*

$$\varphi_h = \frac{\#C_h}{\sum_{h=1}^{b}\#C_h}\cdot\frac{e^{TS_h}}{\sum_{h=1}^{b} e^{TS_h}}. \tag{26}$$

*Consequently, the weighting scheme for subgroup* $C_h$ *can be established as in Equation (27):*

$$\lambda_h = \frac{\varphi_h}{\sum_{h=1}^{b}\varphi_h}. \tag{27}$$

This framework addresses the limitations of conventional approaches by systematically incorporating SN metrics and trust dynamics while maintaining computational tractability. The parametric structure allows decision facilitators to balance between subgroup size importance and quality of social connections, particularly crucial when managing minority opinions [37].

## 4. A CRP Method Based on MOH and DISs

In this section, multigranulation rough set (MGRS) theory is first used to perform consensus measurement, from which the opinion divergence degree within the group is derived. This step accurately assesses the consistency and differences in opinions by considering the evaluations at different granularity levels. Next, for groups with high opinion divergence, minority opinion detection and handling are carried out to ensure that minority opinions are effectively identified and appropriately integrated into the decision-making process. Finally, prospect regret theory is applied for weight adjustment, taking into account psychological factors such as regret and risk preferences of DMs. This ensures that the weight of group members are adjusted in a rational manner, ultimately leading to consensus.

*4.1. Consensus Measurement*

This section provides a detailed explanation of how to use the MGRS theory in consensus measurement. By incorporating the concept of the multi-granularity concept, we are able to analyze data at different levels and scales [50], thereby obtaining more precise and comprehensive consensus results. This method not only considers the information differences at various granularities but also effectively integrates the information from each granularity level, providing a more flexible and detailed analytical framework for consensus decision-making.

To address this issue, this paper proposes an MGRS approach, specifically applying rough set theory to fuzzy decision-making. Rough set theory is designed to extract both deterministic and possible rules from an inconsistent complete system, effectively handling the widespread inconsistencies found in information systems. In rough set decision-making, any indiscernible set Y is described using two precisely definable sets: the upper approximation $\overline{R}$ and the lower approximation $\underline{R}$. These approximations help characterize uncertain information, improving decision-making accuracy and robustness.

Specifically, the DIS leverages MGRS theory to objectively measure consensus by analyzing decision choices at multiple granular levels. This approach replaces subjective threshold settings with a formula-driven framework, which includes three key indicators ($T1$, $T2$, $T3$) derived from the optimistic and pessimistic approximations of the target set $X$. This structured process, combined with the flowchart in Figure 2, aims to illustrate how DIS systematically quantifies conflicts and guides feedback adjustments, avoiding ambiguities in the research. Below is a structured explanation of the involved terms.

**Definition 17.** *Given an arbitrary, non-definable set $X$, this paper introduces the concepts of lower and upper approximations to describe its boundaries. The lower approximation of $X$, denoted as $\underline{R}(X) = \{x|[x]_R \subseteq X\}$, represents the smallest definable subset entirely contained within $X$. Conversely, the upper approximation, denoted as $\overline{R}(X) = \{x|[x]_R \cap X \neq \varnothing\}$, refers to the largest definable set that encompasses $X$ in its entirety. These approximations provide a structured approach to handling uncertainty, enabling a more precise characterization of sets within rough set theory.*

MG analysis involves approaching a problem from multiple perspectives and levels, utilizing various granular structures to approximate a target concept. Building upon this MG framework, we explore the concept of MGRSs, which can be categorized into two primary models:

1. **Optimistic MGRS**—based on the "seeking common ground while preserving differences" strategy, this model defines rough set approximations in a more inclusive manner, allowing for flexibility in classification.
2. **Pessimistic MGRS**—rooted in the "seeking common ground while rejecting differences" approach, this model applies stricter constraints to ensure a higher level of consistency across granular structures.

**Definition 18.** *Let $X$ be a given target set, the upper approximation and lower approximation of $X$ in the optimistic MGRS model are, respectively, defined as Equations (28) and (29):*

$$\sum_{i=1}^{m} \underline{R_i}^{O}(X) = \{[x]_{R_1} \subseteq X \vee [x]_{R_2} \subseteq X \vee \cdots \vee [x]_{R_m} \subseteq X | x \in U\}, \tag{28}$$

$$\sum_{i=1}^{m} \overline{R_i}^{O}(X) = \left(\sum_{i=1}^{m} \underline{R_i}^{O}(X^c)\right)^c. \tag{29}$$

**Definition 19.** *Let $X$ be a given target set, the upper approximation and lower approximation of $X$ in the pessimistic MGRS model are, respectively, defined as Equations (30) and (31):*

$$\sum_{i=1}^{m} \underline{R_i}^{P}(X) = \{[x]_{R_1} \subseteq X \wedge [x]_{R_2} \subseteq X \wedge \cdots \wedge [x]_{R_m} \subseteq X | x \in U\}, \tag{30}$$

$$\sum_{i=1}^{m} \overline{R_i}^{P}(X) = \left(\sum_{i=1}^{m} \underline{R_i}^{P}(X^c)\right)^c. \tag{31}$$

**Definition 20.** *Based on the above four formulas, the definition of the DISs are as in Equations (32)–(34):*

$$T_1 = \left\{ j | \max_{x_j \in X} \left\{ \sum_{i=1}^{m} \underline{R_i}^{O}(x_j) \oplus \sum_{i=1}^{m} \overline{R_i}^{O}(x_j) \right\} \right\}, \tag{32}$$

$$T_2 = \left\{ k | \max_{x_k \in X} \left\{ \sum_{i=1}^{m} \underline{R_i}^{P}(x_k) \oplus \sum_{i=1}^{m} \overline{R_i}^{P}(x_k) \right\} \right\}, \tag{33}$$

$$T_3 = \left\{ l | \max_{x_l \in X} \left\{ \left\{ \sum_{i=1}^{m} \underline{R_i}^{O}(x_l) \oplus \sum_{i=1}^{m} \overline{R_i}^{O}(x_l) \right\} \oplus \left\{ \sum_{i=1}^{m} \underline{R_i}^{P}(x_l) \oplus \sum_{i=1}^{m} \overline{R_i}^{P}(x_l) \right\} \right\} \right\}. \tag{34}$$

**Rule 2.**

(1) If $T_1 \cap T_2 \cap T_3 \neq \varnothing$ holds, then consensus has been reached.
(2) If $T_1 \cap T_2 \cap T_3 = \varnothing$ and $T_1 \cap T_2 \neq \varnothing$ both hold, then consensus partially reach consensus.
(3) If $T_1 \cap T_2 \cap T_3 = \varnothing$ and $T_1 \cap T_2 = \varnothing$ both hold, then consensus has not been reached.

### *4.2. MOH (M1)*

This section addresses the treatment of minority opinions via two key stages: detection and management. The process begins with identifying divergent opinions within the group, followed by implementing strategies to adjust or integrate these viewpoints. By effectively handling minority opinions, the goal is to enhance group consensus while maintaining decision diversity and fairness.

#### 4.2.1. Detection Mechanism

To objectively and efficiently identify minority opinions within subgroups exhibiting low consensus levels, it is assumed that a subgroup $\#C_{h*}$ represents the minority opinion. The subgroup $\#C_{h*}$ should meet the following three characteristics:

(1) The consensus level of $\#C_{h*}$ and its alignment with the group opinion are the lowest.
(2) $\#C_{h*}$ consists of a limited number of DMs. the minority opinion should satisfy the threshold $\#C_{h*} \leq \overline{d}(\overline{d} = [d/b]$, (where [.] is a floor function).
(3) $\#C_{h*}$ should hold a relatively high level of authority and gain recognition and trust from the majority of DMs within the group.

Therefore, to classify the subgroup $\#C_{h*}$ as a minority opinion subgroup, it must simultaneously satisfy both of the following conditions, as shown in Equation (35):

$$C_{h*} = \begin{cases} |C_h| \leq \left\lfloor \dfrac{d}{b} \right\rfloor; \\ TS_h > TS = \dfrac{1}{b} \displaystyle\sum_{i=1}^{b} TS_i. \end{cases} \tag{35}$$

4.2.2. Management Mechanism

Based on the aforementioned criteria, the subgroup $\#C_{h*}$ is classified as reliable and trustworthy if its opinions are deemed valuable. In such cases, a weight-adjustment mechanism is activated. Conversely, if minority opinions lack credibility or fail to gain support from other DMs, they may impede consensus formation. For such scenarios, Section 4.3 outlines principles to recalibrate the minority subgroup's initial opinions. When $\#C_{h*}$'s opinions are validated as reasonable and credible, their influence is amplified via a trust-driven weight-adjustment process.

**Definition 21.** *Let $\lambda^t = \{{\lambda_1}^t, {\lambda_2}^t, \cdots, {\lambda_b}^t\}$ denote the subgroup weight vector at iteration t. For subgroup $C_{h*}$ holding minority opinions in round t, its adjusted weight ${\lambda_{h*}}^{t+1}$ for the subgroup $C_{h*}$, is computed as in Equation (36):*

$$ {\lambda_{h*}}^{t+1} = \frac{\overline{TS}}{TS_{h*}} \cdot {\lambda_{h*}}^t + (1 - \frac{\overline{TS}}{TS_{h*}}) \cdot {\lambda_{\max}}^t, \tag{36} $$

*where $\lambda_{\max}^t$ denotes the maximum weight of any subgroup in the t-th round. Once the adjusted weights for the minority opinion subgroup have been calculated, it is necessary to normalize all subgroup weights to ensure that the sum of the normalized weights equals 1, i.e., $\sum_{h=1}^{b} \lambda_h^{t+1} = 1$, This normalization step ensures consistency in the weight distribution and maintains the integrity of the decision process.*

This design addresses a common issue in classical methods, where weight adjustments can escalate drastically, sometimes expanding by tens of times. In these cases, the changes may go unrecognized by other subgroups, leading to unreasonable outcomes. By constraining weight adjustments in this way, the proposed method ensures a more balanced and rational decision process.

*4.3. Opinion Divergence-Driven Dynamic Weight Adjustment (M2)*

For subgroups that consistently show the lowest consensus levels under the M1 method and remain unaddressed, these clusters are often considered impediments to effective decision-making. In this section, we introduce targeted modifications designed to enhance the CRP by directly addressing these problematic subgroups.

Traditional consensus feedback mechanisms tend to presuppose that DMs' opinions are broadly acceptable. In practice, however, psychological factors play a significant role in how individuals update their views, often leading to overly rationalized decisions that lean heavily on subjective experience. As a result, final outcomes may not adequately capture the objective realities of the situation.

To overcome this limitation, we propose a novel framework within the context of SNs. This framework integrates prospect–regret theory with an analysis of interpersonal influences, thereby mitigating the bias introduced by psychological factors. By contrasting the original evaluations with an ideal group assessment, our approach not only reduces subjectivity but also anchors feedback in a more objective baseline. Moreover, by incorporating the dynamics of social relationships-recognizing that DMs frequently consult peers-the modified feedback mechanism better mirrors real-world decision processes.

The subsequent sections provide a detailed exposition of these modification rules, outlining how they contribute to more robust and realistic decision outcomes.

**Step 1. The endogenous reference point generation**

(1) The subgroup reference point definition using Equation (37):

$$\kappa_{ij}^{h} = \frac{1}{n}\sum_{k=1}^{n}\frac{\mu_{ik}^{h} + \nu_{ik}^{h}}{2}. \tag{37}$$

- Note: Dynamically generated from subgroup historical evaluations.

(2) The prospect-deviation calculation using Equation (38):

$$\sigma_{ij}^{h} = \left|\frac{\mu_{ij}^{h} + \nu_{ij}^{h}}{2} - \kappa_{ij}^{h}\right|. \tag{38}$$

- Function: Quantifies deviation from reference standards.

**Step 2. The prospect–regret joint valuation**

(1) The nonlinear value function using Equation (14):

$$v_{ij}^{h} = \begin{cases} (\kappa_{ij}^{h} - \sigma_{ij}^{h})^{\alpha}, & \text{gain domain} \\ -\lambda(\sigma_{ij}^{h} - \kappa_{ij}^{h})^{\beta}, & \text{loss domain} \end{cases}. \tag{39}$$

(2) The social comparison regret value using Equation (15) :

$$R_{ij}^{h} = \sum_{e \neq h}\left(1 - e^{-\delta(\kappa_{ij}^{e} - \kappa_{ij}^{h})}\right). \tag{40}$$

(3) The integrated prospect value using Equation (41):

$$V^{h} = \frac{1}{n \times m}\sum_{i=1}^{n}\sum_{j=1}^{m}(v_{ij}^{h} - R_{ij}^{h}). \tag{41}$$

**Step 3. Subgroup weight iteration**

(1) The dynamic weight update using Equation (42):

$$\lambda_{h}^{t+1} = \lambda_{h}^{t} + \gamma \cdot (1 - V^{h}) \cdot \lambda_{h}^{t}. \tag{42}$$

- Parameter: $\gamma \in [0, 1]$ (recommended initial value: 0.2).

(2) The weight normalization using Equation (43):

$$\lambda_{h}^{t+1} = \frac{\lambda_{h}^{t+1}}{\sum_{k=1}^{b}\lambda_{k}^{t+1}}. \tag{43}$$

The parameter $\gamma = 0.2$ was selected to balance convergence speed and stability. A smaller $\gamma$ (e.g., $\gamma < 0.2$) risks overly sluggish adjustments, while a larger $\gamma$ (e.g., $\gamma > 0.3$) may cause oscillations or premature convergence. Therefore, set the initial value to 0.2.

**Step 4. Consensus target matrix update** The target matrix generation using Equations (44) and (45):

$$\mu_{j}{}^{B} = \frac{1}{b \times n \times m}\sum_{h=1}^{b}\sum_{i=1}^{n}\sum_{j=1}^{m}\mu_{ij}{}^{b}\lambda_{h}, \tag{44}$$

$$v_{j}{}^{B} = \frac{1}{b \times n \times m}\sum_{h=1}^{b}\sum_{i=1}^{n}\sum_{j=1}^{m}v_{ij}{}^{b}\lambda_{h}. \tag{45}$$

The pseudocode meticulously outlining the full operational procedure of the CRP is vividly showcased in Algorithm 2.

**Algorithm 2** A CRP method based on DISs and MOH

**Input:** The subgroup partition $\{C_h\}$, the initial weights $\{\lambda_h^0\}$, the initial target matrix $B$.
**Output:** The final ranking $X_{\text{final}}$.

**Step 1: The consensus measurement**
**repeat**
  Compute MGRS indicators: $T_1$, $T_2$, $T_3$ using Equations (32)–(34);
**until** convergence condition met;
**Step 2: Minority opinion handling**
**if** $T_1 \cap T_2 \cap T_3 = \emptyset$ **then**
  Identify the candidate minority subgroups $C_{h^*}$ using Equation (35);
  **for** each $C_{h^*}$ **do**
    **if** credibility verification passes **then**
      Adjust subgroup weights using Equation (36);
    **else**
      Start **Step 3**;
    **end if**
  **end for**
**end if**
**Step 3: Dynamic weight update**
Compute the subgroup foreground-regret value $V^h$ using Equation (41);
Update the weight $\lambda_h^{t+1}$ using Equation (42);
Normalize weights using Equation (43);
**Step 4: Target matrix evolution**
Update the consensus matrix $B$ using Equations (44) and (45);
**Step 5: Generate results**
Acquire the comprehensive score $SC_{ij}$ from $T_3$;
**return** ranking the result $X_{\text{final}}$.

## 5. Empirical Validation and Analysis

Case study experiments are presented in this section to assess and analyze the performance of the proposed LSGDM modeling framework. We first provide a comprehensive explanation of how the proposed method can be applied to real-world decision-making problems. Furthermore, we conduct both comparative and sensitivity analyses to evaluate how variations in model parameters influence its stability, thereby validating its feasibility and effectiveness. Finally, we discuss the practical implications of the proposed model in managing minority group opinions, highlighting its potential to enhance decision-making processes.

### *5.1. Empirical Validation*

In this section, we collect online reviews of drone products from 13 verified stores on the JD.com platform (https://www.jd.com/ (16 January 2025)), a leading Chinese e-commerce platform renowned for its comprehensive and authentic product evaluations. Spanning from January 2023 to December 2024, the dataset encompasses two years of consumer feedback, strategically curated to capture temporal trends in UAV preferences and ensure representativeness. A total of 1280 reviews were systematically curated, corresponding to five UAV models evaluated across five key attributes: operability, functionality, appearance, durability, and quality. These attributes were selected to assist users in making informed purchasing decisions. The following will illustrate the application of the model via the experimental process.

**Step 1: Data fuzzy processing**

For textual evaluations from 13 suppliers on five critical UAV attributes (e.g., operation, functionality), sentiment scores $s_{ij} \in [-1, 1]$ are computed, then transformed into IFNs $(\mu_{ij}, v_{ij})$. The complete DM opinion matrix is provided in Appendix A.

**Step 2: Completing the trust propagation matrix**

The trust matrix among different merchants is completed via propagation using Equation (13), resulting in a fully completed trust matrix. Detailed information about this trust matrix can be found in Appendix B. Figure 3 shows the heatmaps of the incomplete trust matrix and the completed trust matrix.

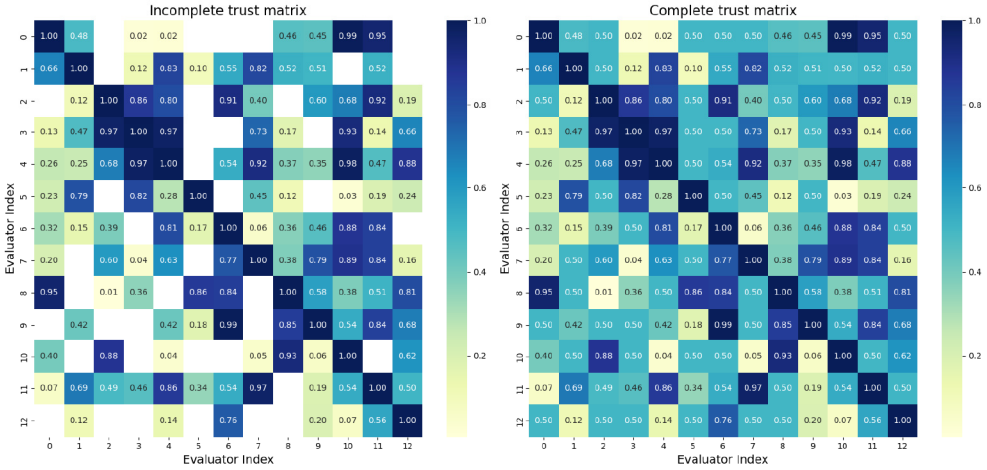



**Figure 3.** Comparison of heatmaps: incomplete vs completed trust matrices.

**Step 3: Clustering and aggregation processes**

The elbow criterion was employed to determine the optimal cluster number $k$ via curvature analysis of the SSE trajectory. For our cohort of thirteen DMs, the concavity variation rate reached statistical significance at $k = 4$, indicating an optimal trade-off between model complexity and information loss (Figure 4). Next, the DMs are grouped using the enhanced K-means++ clustering method, as shown in Table 4. Afterward, aggregation is performed within each subgroup of DMs to obtain the evaluation matrix for the subgroups, with the detailed information provided in Appendix C.

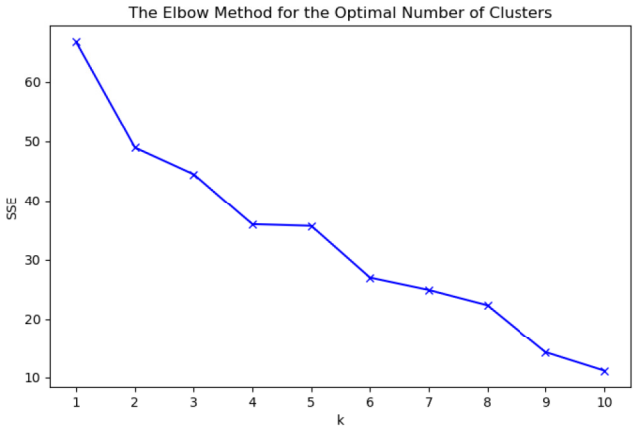



**Figure 4.** Elbow method: SSE vs. number of clusters ($k$).

**Table 4.** Result of enhanced K-means++ clustering analysis.

| The Grouping of DMs |
|---|
| $G_0 = \{e_3, e_5, e_7, e_9, e_{10}, e_{12}\}$ |
| $G_1 = \{e_8\}$ |
| $G_2 = \{e_6, e_{13}\}$ |
| $G_3 = \{e_1, e_2, e_4, e_{11}\}$ |

**Step 4: Calculating the DMs and group weights of different DMs**

The weights of each subgroup are determined based on two key factors: the closeness of trust relationships among members within the group and the size of the group's membership. Specifically, subgroups with stronger internal trust connections and larger membership sizes are assigned proportionally greater weights.

Additionally, trust correlations between different subgroups are calculated using trust relationships identified within each subgroup. The final computational results are presented in Tables 5 and 6, where all data are retained to three decimal places for precision.

**Table 5.** Weights of DMs.

| Weights of DMs |
|---|
| 0.062 0.065 0.081 0.072 0.079<br>0.067 0.096 0.080 0.072 0.067<br>0.091 0.090 0.078 |

**Table 6.** Groups' weights.

| $G_h$ | $\#C_h$ | $TS_h$ | $\lambda_h$ |
|---|---|---|---|
| G0 | 6 | 1.889 | 0.044 |
| G1 | 1 | 6.612 | 0.833 |
| G2 | 2 | 3.606 | 0.082 |
| G3 | 4 | 2.204 | 0.041 |

**Step 5: The consensus measurement**

A multi-criteria decision matrix is utilized to quantify inter-subgroup consensus levels, with an initial target preference matrix

$$B = \Big[(0.8, 0.2) \quad (0.5, 0.5) \quad (0.2, 0.8) \quad (0.9, 0.1) \quad (0.7, 0.3)\Big].$$

Three-phase consensus indicators ($T_1$–$T_3$ ) are computed to validate group convergence using Equations (29) and (31)–(34) (see Table 7 for experimental results).

**Table 7.** DIS outcomes.

| Indexes | Values | The Greatest Subset |
|---|---|---|
| $T1$ | $[(0, 0.259), (1, 0.262), (2, 0.21), (3, 0.237), (4, 0.277)]$ | $\{4\}$ |
| $T2$ | $[(0, 0.118), (1, 0.123), (2, 0.054), (3, 0.184), (4, 0.168)]$ | $\{3\}$ |
| $T3$ | $[(0, 0.377), (1, 0.385), (2, 0.264), (3, 0.421), (4, 0.444)]$ | $\{4\}$ |

With the subgroup number and its corresponding score value, and taking the subgroup corresponding to the largest score value, we obtain $T1 \cap T2 \cap T3 = \varnothing$, but $T1 \cap T3 \neq \varnothing$, so there is a partial contradiction.

**Step 6: CRP**

(1) Round 1

After testing, it is concluded using Equation (35) that there are minorities in subgroup 1 and subgroup 2. The management mechanism is implemented for the groups with minorities, the weights are adjusted by Equation (36), and the final weights of the subgroups are obtained as shown in Table 8. The target matrix $B$ is updated by using the weights. After obtaining the new target matrix B, the degree of consensus is re-measured.

**Table 8.** Adjusted group weight.

| $G_0$ | $G_1$ | $G_2$ | $G_3$ |
|---|---|---|---|
| 0.057 | 0.070 | 0.603 | 0.270 |

After re-measuring for consensus, the values of $T1$, $T2$, and $T3$ are presented in Table 9. Since there is an intersection among these values, a consensus is reached, terminating the current loop.

**Table 9.** Adjusted DIS outcomes.

| Indexs | Values | The Greatest Subset |
|---|---|---|
| $T1$ | $[(0,0.411),(1,0.414),(2,0.4),(3,0.406),(4,0.414)]$ | {1,4} |
| $T2$ | $[(0,0.406),(1,0.414),(2,0.375),(3,0.392),(4,0.414)]$ | {1,4} |
| $T3$ | $[(0,0.817),(1,0.828),(2,0.775),(3,0.798),(4,0.828)]$ | {1,4} |

**Step 7: Reaching the final result**

The final score of each alternative plan is determined using the weighted average of different attribute scores. A line graph is used to visually represent these scores (when the score values are equal, the group is ranked in order). The result for the case is presented in Table 10.

**Table 10.** Ranking results.

| Ranking Results |
|---|
| $x_0 \succ x_1 \succ x_4 \succ x_3 \succ x_2$ |

### 5.2. Sensitivity Analysis

Within the framework of prospect–regret theory, this section presents sensitivity tests on parameters $\alpha$, $\beta$, and $\lambda$ to deeply analyze the specific effects of their fluctuations on the ranking of solutions.

Figure 5 presents radar charts of rankings corresponding to different $\alpha$ values, with $\beta$ fixed at 0.92 and $\lambda$ at 2.25. Visually, despite variations in $\alpha$, the relative positions of the solutions ($x_0$- $x_4$) on the radar chart almost overlap, with no significant differences noticeable. This strongly suggests that under the given conditions, changes in $\alpha$ have negligible impact on the ranking of solutions, with the results exhibiting high stability.

A similar trend is observed when comparing Figures 6 and 7. In Figure 6, with $\alpha$ and $\lambda$ fixed, changes in $\beta$ do not lead to any noticeable dispersions or reordering of the solutions on the radar chart. The relative positions of the solutions remain stable, suggesting that the $\beta$ parameter has minimal disturbance on the ranking. In Figure 7, after fixing $\alpha$ and $\beta$, variations in $\lambda$ also do not significantly affect the ranking, and the overall shape of the radar chart and the relative positions of the solutions remain consistent.

Considering the sensitivity test results for these three parameters, we can conclude that under the experimental conditions set in this study, changes in $\alpha$, $\beta$, and $\lambda$ do not have a significant or substantial effect on the solution ranking. This suggests that the theoretical model exhibits high stability and robustness within the current parameter range.

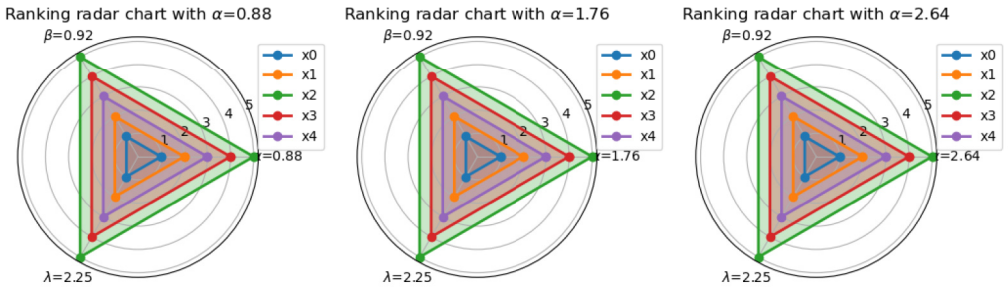



**Figure 5.** Ranking changes with $\alpha$.

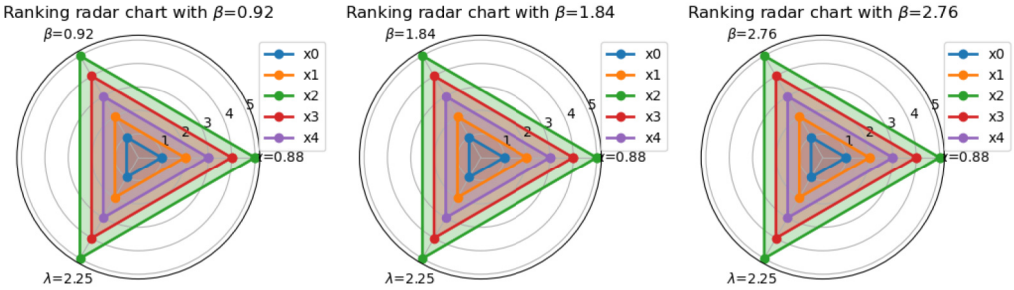



**Figure 6.** Ranking changes with $\beta$.

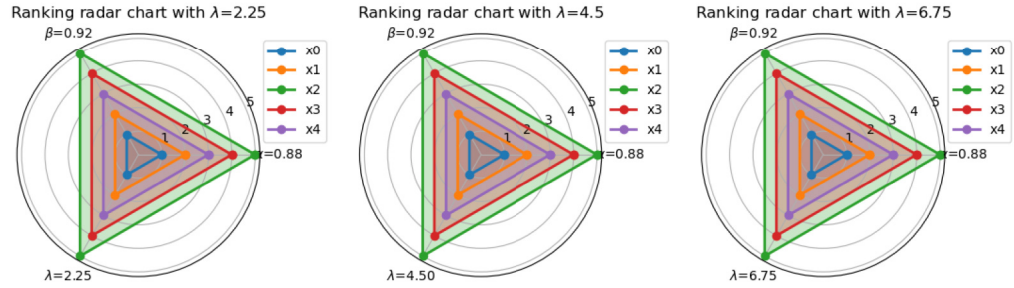



**Figure 7.** Ranking changes with $\lambda$.

### *5.3. Simulation Analysis*

In this section, we present a series of simulation tests designed to evaluate the feasibility of the proposed method on a moderate scale. All simulations were carried out using PyCharm Community Edition 2024.3.3 on a personal computer equipped with an Intel Core i5-12500H processor running at 2.50 GHz. These tests aim to examine how variations in the number of DMs and the scale of the decision problem affect the overall CRP time.

Figure 8 is a visualization of the time distribution for the LSGDM process under various decision scales ($d = 20, 100, 200, 400, 800$, and 1000) and different problem sizes ($n$ and $m$ ranging from 3 to 7). The horizontal axis represents the trial number (from 0 to 100), while the vertical axis shows the time required (in seconds) for each trial. Each colored line and marker corresponds to a specific $(n, m)$ combination.

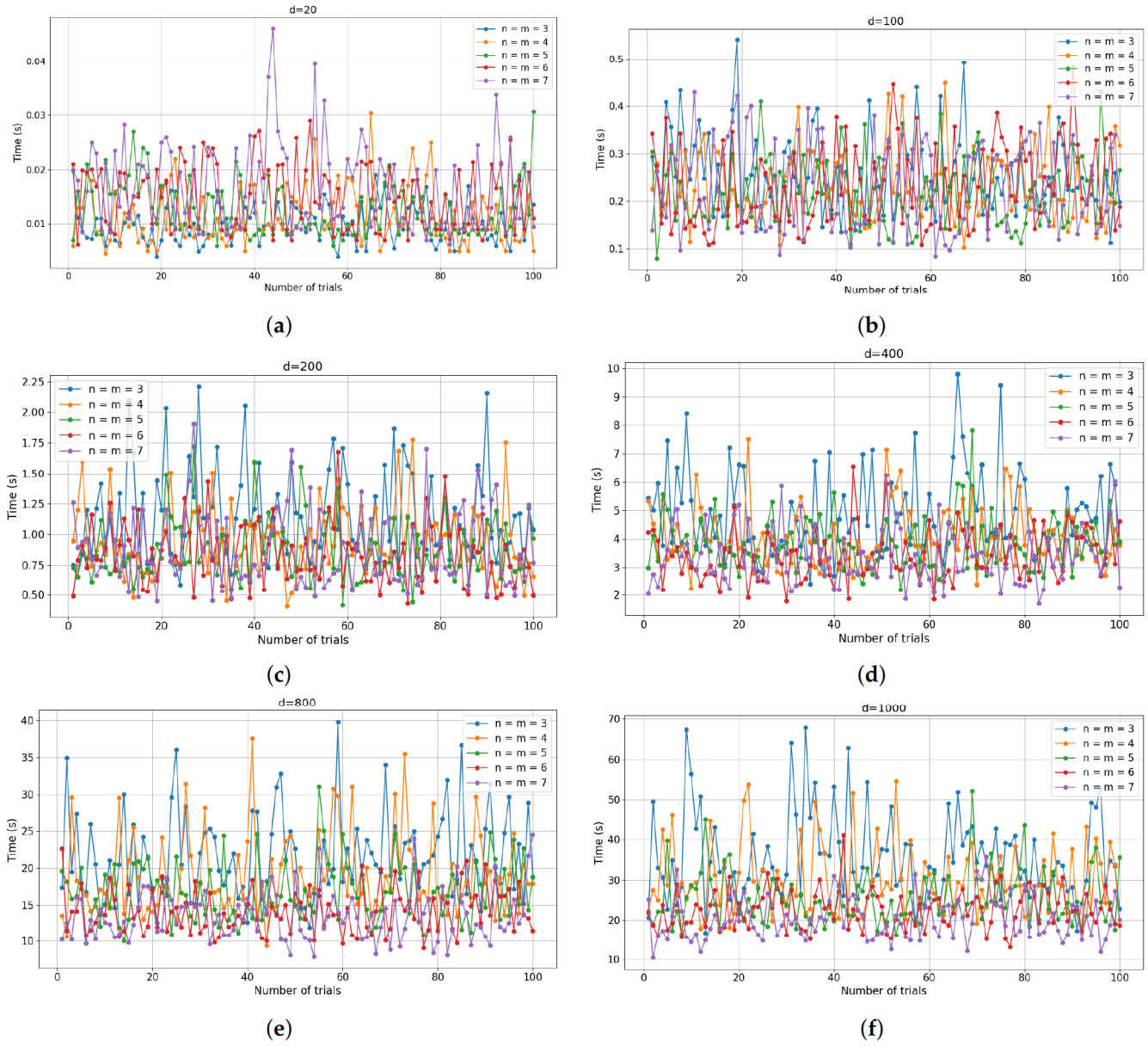



**Figure 8.** Performance analysis of execution time across various parameter settings.

From Figure 8, it is evident that as $d$ (the number of DMs) increases, the overall running time tends to rise. In particular, when $d$ reaches the hundreds or thousands, certain $(n, m)$ combinations exhibit significant spikes in time consumption. For instance, at $d = 800$ or $d = 1000$, some curves show notably higher peak values compared to smaller-scale scenarios. This indicates that when both the decision scale and problem dimension grow, the computational complexity of the LSGDM process increases, resulting in more pronounced time overhead.

Furthermore, for a fixed $d$ value, different $(n, m)$ combinations lead to varying time distributions. Some combinations (e.g., $(n = 3, m = 3)$ or $(n = 4, m = 4)$) generally require less time, with relatively minor fluctuations across most trials. By contrast, when both $n$ and $m$ are larger (e.g., $(n = 7, m = 7)$), there are higher time-consumption peaks during multiple trials, indicating that higher-dimensional problems demand greater computational effort and exhibit less predictable run times.

Overall, these plots illustrate that as both the decision scale $d$ and the problem dimensions $(n, m)$ increase, the running time for the LSGDM method shows a greater variability

and a clear upward trend. These findings offer insights for future research focused on algorithmic optimization, parallelization, or improved data structures.

### *5.4. Comparison Analysis*

To further substantiate the superiority and practicality of the method put forward in this paper, this section meticulously conducts comparative experiments against a set of well-established existing methods. These experiments are designed to systematically evaluate how the proposed method performs relative to others in terms of crucial aspects such as execution efficiency, consensus-reaching effectiveness, and the rationality of sorting results, thereby offering a comprehensive and reliable validation of its advantages.

#### 5.4.1. Component Necessity Analysis via Ablation Study

To comprehensively evaluate the functional necessity of each algorithm component, we mainly conducted systematic ablation experiments across three dimensions: computational efficiency, consensus formation quality, and ranking effectiveness. By applying a control variable approach, we systematically removed each sub-method and observed its impact on specific process indicators.

Table 11 summarizes the deformation methods implemented in our proposed framework.

**Table 11.** A summary of the required comparison methods.

| | |
|---|---|
| Our method | $m0$ |
| Our method eliminates the ranking relationship. | $m1$ |
| Our method eliminates MOH | $m2$ |
| Our method replaces DIS with weighted sum during decision-making | $m3$ |
| Our method eliminates prospect–regret theory | $m4$ |

A benchmarking framework employing Monte Carlo iterations ($n = 100$) was implemented to quantitatively assess performance differentials between baseline method $m0$ and enhanced method $m1$, with a particular focus on algorithmic refinements in cluster optimization. Statistical aggregates of these metrics, calculated over 100 independent trials, are systematically tabulated in Table 12 (values truncated at 3 decimal places for precision consistency).

**Table 12.** An in-depth analysis of sub-methods in clustering.

| Methods | Total Execution Time | Clustering Time | Initial Consensus Degree | Final Ranking |
|---|---|---|---|---|
| $m0$ | 0.0234 s | 0.0188 s | Partial contradiction | $x_0 \succ x_1 \succ x_4 \succ x_3 \succ x_2$ |
| $m1$ | 0.0091 s | 0.0045 s | Partial contradiction | $x_1 \succ x_4 \succ x_0 \succ x_3 \succ x_2$ |

Based on the comparative analysis of methods $m0$ and $m1$, it can be concluded that method $m0$ exhibits certain advantages. In terms of the total execution time, although $m0$ takes 0.0234 s which is longer than $m1$'s 0.0091 s, when considering the clustering execution time alone, the difference between $m0$'s 0.0188 s and $m1$'s 0.0045 s is not extremely significant. More importantly, in terms of the initial consensus degree, both methods show partial contradiction, but method $m0$ may have a more stable underlying consensus-reaching mechanism during the clustering process. Regarding the final sorting result, the ranking sequence provided by $m0$ might be more in line with the expected decision-making logic or practical application requirements. Overall, considering multiple key metrics comprehensively, method $m0$ is a relatively better choice in this clustering-related scenario.

In an endeavor to discern how the inclusion of MOH impacts the outcomes within the CRP, a meticulous comparative analysis is conducted. This analysis juxtaposes scenarios featuring MOH against those lacking it. Three pivotal metrics serve as the cornerstone of this comparison: The CRP time, the number of iterations, the final sorting result.

To ensure the reliability of the results, both conditions, with and without MOH, are executed a hundred times each. During these runs, corresponding data for each metric are carefully recorded. Subsequently, the average values for every indicator are calculated and presented. All these comprehensive findings are neatly summarized in Table 13, offering a clear-cut view of the necessity of MOH in CRP.

**Table 13.** Analysis of the necessity of MOH in CRP.

| Methods | CRP Time | Number of Iterations | Final Sorting Result |
|---|---|---|---|
| $m0$ | 0.0046 s | 1 | $x_0 \succ x_1 \succ x_4 \succ x_3 \succ x_2$ |
| $m2$ | 0.0012 s | 1 | $x_4 \succ x_1 \succ x_0 \succ x_3 \succ x_2$ |

Based on the comparative analysis of scenarios with and without MOH in the CRP, as presented in Table 13, in terms of CRP time, although $m2$ is quicker with a time of 0.0012 s compared to $m0$'s 0.0046 s, the difference is relatively small. Significantly, both methods have the same number of iterations, indicating that $m0$ does not sacrifice efficiency in terms of the iterative process. More importantly, the final sorting result provided by $m0$ might align better with the anticipated decision-making logic or practical application requirements. Considering these pivotal metrics comprehensively, $m0$ proves to be a more suitable choice, highlighting the relative necessity and advantage of its approach in the CRP context.

An in-depth comparative analysis is conducted to explore the implications of substituting the DIS with a weighted sum during the decision-making process, as presented in Table 14. This analysis focuses on two critical metrics: the CRP time and the final sorting result. By juxtaposing method $m0$ against $m3$, this analysis aims to gain a comprehensive understanding of the impact of this substitution.

**Table 14.** Analysis of replacing DISs with weighted sum during decision-making.

| Methods | The CRP Time | The Final Sorting Result |
|---|---|---|
| $m0$ | 0.0046 s | $x_0 \succ x_1 \succ x_4 \succ x_3 \succ x_2$ |
| $m3$ | 0.0054 s | $x_1 \succ x_4 \succ x_0 \succ x_3 \succ x_2$ |

The comparative analysis in Table 14 examines the effect of replacing DIS with a weighted sum in the decision-making context. Focusing on CRP time and the final sorting result, it contrasts method $m0$ and $m3$. $m0$ shows a CRP time of 0.0046 s, shorter than $m3$'s 0.0054 s. Meanwhile, both methods result in the same final sorting order. Efficiency in decision-making is often crucial, and with a shorter CRP time while achieving an identical result, $m0$ proves to be more advantageous. This comparison offers a clear view of the performance differences between the two methods, which can guide the selection of appropriate decision-making approaches in practical applications.

To evaluate the impact of integrating prospect–regret theory into the CRP, we conduct a comparative experiment between two scenarios: CRP with prospect–regret theory ($m0$) and CRP without prospect–regret theory ($m4$). The former employs the prospect–regret theory-based weight-adjustment mechanism, while the latter uses a uniform weight update rule without considering risk preferences or regret aversion. Key metrics include CRP time, number of iterations, and the rationality of the final ranking result. Using the same initial

trust matrix and clustering results. Each condition is executed 100 times to ensure statistical reliability, with average values reported in Table 15.

**Table 15.** Comparison of CRP with and without prospect–regret theory.

| Methods | CRP Time | Number of Iterations | Final Sorting Result |
|---|---|---|---|
| $m0$ | 0.0046 s | 1 | $x_0 \succ x_1 \succ x_4 \succ x_3 \succ x_2$ |
| $m4$ | 0.0032 s | 4 | $x_1 \succ x_4 \succ x_0 \succ x_3 \succ x_2$ |

$m0$ takes slightly longer (0.0046 s) than $m$ 4 (0.0032 s) primarily due to the added computational overhead of the nonlinear value function and regret calculation in prospect–regret theory. However, the time difference is minimal, suggesting that the inclusion of behavioral factors does not significantly impact efficiency. Notably, $m0$ reaches consensus in just one iteration, whereas $m4$ requires four iterations. This highlights the advantage of prospect–regret theory in accelerating convergence by directing weight adjustments toward a risk-sensitive consensus, thereby reducing the need for multiple iterations. Overall, the results demonstrate that incorporating behavioral factors enhances convergence speed with only a marginal computational cost.

#### 5.4.2. Comparative Analysis of Trust Propagation Operators

To validate the superiority of our trust propagation mechanism, we compare the Hamacher-based operator (denoted as $M0$) with two benchmark methods across three critical metrics: computational efficiency, error control (relative error), and the ability to model trust decay in SNs. Table 16 presents the representations and descriptions of the three methods.

For the three operators, we randomly generated initial trust matrices with some empty entries. After completing them through trust propagation, we conducted 100 experiments on indicators like relative error and took the averages, which are summarized in Table 17 to compare the operators' performance.

**Table 16.** Method descriptions.

| | | |
|---|---|---|
| Our method | $M0$ | Our proposed method, which integrates the Hamacher aggregation operator with a decay factor to dynamically refine trust values, accounting for both multiplicative interactions and distance-dependent decay in SNs. |
| Weighted averaging | $M1$ | A classical approach that estimates missing trust values using weighted averages of direct and indirect trust relationships, without explicit decay modeling. |
| Heuristic operators | $M2$ | A heuristic-based method that prioritizes short propagation paths but lacks a systematic framework for decay factor adjustment. |

**Table 17.** A qualitative analysis comparing the trust propagation operators.

| Methods | Relative Error | Average Time/Iteration (s) | Decay Modeling Support |
|---|---|---|---|
| $M0$ | 0.9992 | 0.00124 | Yes |
| $M1$ | 1.3994 | 0.00005 | No |
| $M2$ | 2.5137 | 0.00153 | No |

$M0$ achieves the lowest relative error (0.9992), demonstrating superior accuracy in reconstructing incomplete trust matrices. This is attributed to the Hamacher operator's

ability to balance conjunctive and disjunctive trust interactions, unlike Methods A and B, which either oversimplify relationships ($M1$) or ignore decay dynamics ($M2$).

Only $M0$ explicitly incorporates a decay factor to mitigate trust inflation over long paths, a critical feature for realistic SN modeling. While $M1$ is marginally faster, it lacks decay modeling; $M0$'s moderate time overhead (0.001242 s/iteration) is justified by its superior behavioral fidelity and error control, essential for complex LSGDM scenarios.

This comparison highlights that our dual strategy of Hamacher aggregation and decay factor adjustment effectively balances mathematical rigor, behavioral realism, and computational practicality in SN-LSGDM, outperforming classical and heuristic alternatives in both accuracy and theoretical consistency.

#### 5.4.3. Comparison Among Different Clustering Methods

In clustering analysis, the selection of distance metric is pivotal in dictating the clustering results. Classical single-distance clustering methods, such as those widely used and well-established approaches, construct their clustering models relying on distinct distance measures. This leads to significant disparities in their efficacy and applicability across various scenarios.

To offer a clear and comprehensive juxtaposition of these classical methods, we meticulously compiled Figure 9. This figure contrasts our single-distance clustering approach with the classical ones proposed by Huang et al. [41] and Chen et al. [29]. The comparison spans four crucial dimensions: the name of the method, the components of the distance metric, the clustering time and the average subgroup credibility.

Based on the comparative analysis of single-distance clustering methods in Table 14, in terms of clustering time, our method only takes 0.0188 s, which is more efficient than Huang et al. [41]'s 0.0582 s and the simple Euclidean distance-based method's 0.1872 s. Although Cheng et al. [29]'s method has a shorter time of 0.0075 s, our method shows greater advantages in other aspects.

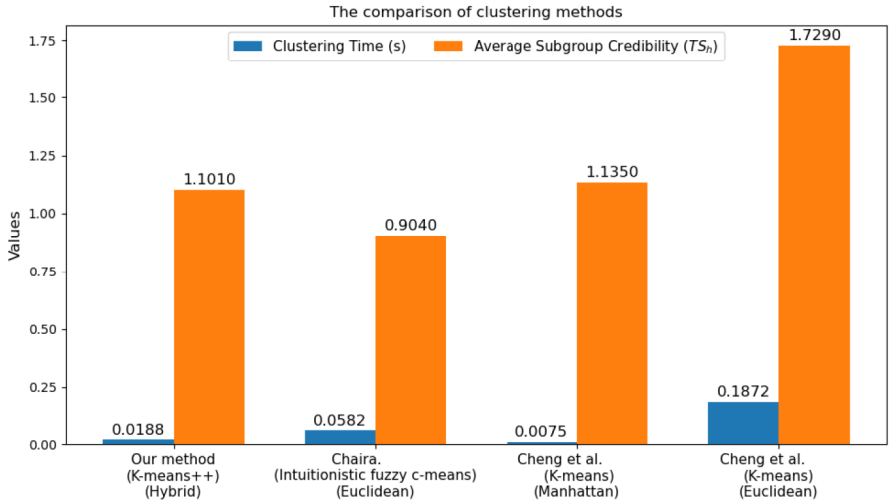



**Figure 9.** Comparison of clustering methods [30,44].

Regarding the average subgroup credibility, our method reaches 1.101, which is higher than Huang et al. [41]'s 0.904. This indicates that the subgroups formed by our method have higher credibility and better internal consistency. Additionally, our method combines Hamming distances based on ranking results, Euclidean distances, and Manhattan distances, enabling it to comprehensively consider different characteristics of data. This multi-distance-metric approach endows our method with stronger adaptability to various data distributions and structures, making it more suitable for complex clustering tasks.

5.4.4. Comparison Between the Proposed Method and Existing Consensus Models

To enable a systematic comparison, existing models are categorized into three groups: traditional LSGDM models (TL), enhanced LSGDM models integrated with SN (SNL), and behavior-management-oriented LSGDM models to handle misbehaviors (ML). This classification framework allows for a clearer presentation of the applicability scenarios and performance variations among different consensus models.

Yuan et al. [51] developed a fuzzy LSGDM method using trust relationships and opinion similarities, applying the improved Louvain algorithm for clustering and weight determination. A two-stage dual fine-tuning CRP reduced decision-making scale, and prospect–regret theory ranked the schemes.

Chen et al. [52] developed an emergency decision-making method for large groups based on uncertain linguistic cloud similarity, which transforms linguistic preferences into cloud models, clusters DMs, determines weights, and reaches consensus via trust updates and opinion interactions.

Meng et al. [53] addressed the shortcomings of consensus-adjustment mechanisms in LSGDM. They introduced cooperative game theory, proposed two-stage Shapley and core-Nash bargaining consensus-adjustment mechanisms.

Liang et al. [54] constructed a threshold-free model for LSGDM based on SNA. The model involves adjusting the trust degree, conducting comprehensive clustering, using the SNDG model to adjust preferences, terminating the feedback based on the consensus conversion rate.

Liu et al. [19] proposes a consensus model for LSGDM driven by an SN. It improves the network through a two-stage analysis, determines expert weights, and constructs a consensus optimization model to achieve minimum adjustments and promote CRP.

Liu et al. [55] developed a consensus model for managing non-cooperative behavior in LSGDM under a dynamic dual SNs. This model constructs a dynamic dual trust network and designs a minimal adjustment consensus model and non-cooperative behavior-management method based on it.

Xu et al. [39] presents an improved consensus model for multi-criteria large-group emergency decision-making, which clusters and analyzes group opinions, determines parameters such as the comprehensive adjustment coefficient, and addresses minority opinions and non-cooperative behaviors to reach a consensus.

Xu et al. [11] proposed the HOAL-MOH-DFT model. It optimizes Leiden clustering with the hippopotamus optimization algorithm, groups based on trust relationships and opinion similarities, and conducts two-stage adjustments to handle various situations in group decision-making. Table 16 compares the core features of the consensus model proposed in this study with above methods, emphasizing the innovative strengths of our approach.

The comparative analysis in Table 18 reveals significant limitations in the key mechanism designs of existing LSGDM models:

TL, while simplifying processes via basic clustering, relies on single indicators (e.g., Chen et al. [52] and Meng et al. [53]) to capture the multidimensional correlations of DMs' behaviors, resulting in incomplete behavioral characterization. Additionally, their lack of a quantitative framework for consensus level (DIS) and proactive intervention in MOH often leads to majority-dominated decision bias.

SNL models enhance opinion adjustment by integrating SN trust relationships, yet their static SNs struggle to adapt to dynamic scenarios. Furthermore, the decoupling of DIS from decision-generation fragments consensus management. ML models attempt to address minority opinions but fail to balance group diversity and decision efficiency due to reliance on single clustering indicators.

**Table 18.** A comparative analysis of consensus methods from a qualitative approach.

| Types | References | Trust Propagation | Clustering Based on Hybrid Distance | DIS | MOH | Managing Opinions with Reference to Trust |
|---|---|---|---|---|---|---|
| TL | Yuan et al. [51] | No | No | No | Yes | No |
| | Chen et al. [52] | No | No | No | No | Yes |
| | Meng et al. [53] | No | No | No | No | No |
| SNL | Liang et al. [54] | No | No | No | Yes | Yes |
| | Liu et al. [19] | No | No | No | Yes | Yes |
| ML | Liu et al. [55] | Yes | No | Yes | Yes | Yes |
| | Xu et al. [39] | Yes | No | Yes | Yes | No |
| | Xu et al. [11] | No | No | Yes | Yes | No |
| | Our method | Yes | Yes | Yes | Yes | Yes |

In contrast, the proposed method overcomes these bottlenecks via multidimensional collaborative design: The integration of dynamic trust propagation and hybrid distance clustering resolves group partitioning bias; the DIS establishes a closed-loop linkage between objective consensus measurement (e.g., entropy-weighted deviation) and decision generation, eliminating the "measurement-decision decoupling" flaw in traditional models; simultaneously, the dynamic weighting strategy for MOH and hierarchical behavior-management mechanism ensure iterative integration of minority opinions while precisely suppressing both explicit and implicit misbehaviors. This organic synergy among mechanisms significantly enhances the model's complexity, inclusivity, and practicality, providing a more realistic solution for LSGDM.

The above approach offers a qualitative comparison between existing methods and the proposed LSGDM approach. To facilitate a more objective comparison and present more convincing results, Table 19 provides a quantitative analysis comparing the LSGDM method with other methods from machine learning.

**Table 19.** A quantitative analysis comparing our method with other methods.

| Methods | Total Time (s) | Iterations | Final Ranking | Ranking Stability |
|---|---|---|---|---|
| Liang et al. [54] | 0.0350 | 3 | $x_1 \succ x_0 \succ x_4 \succ x_3 \succ x_2$ | 0.880 |
| Liu et al. [55] | 0.0089 | 3 | $x_1 \succ x_0 \succ x_4 \succ x_3 \succ x_2$ | 0.892 |
| Xu et al. [39] | 0.0189 | 2 | $x_0 \succ x_1 \succ x_4 \succ x_3 \succ x_2$ | 0.915 |
| Xu et al. [11] | 0.0458 | 5 | $x_4 \succ x_1 \succ x_0 \succ x_3 \succ x_2$ | 0.827 |
| Our method | 0.0234 | 1 | $x_0 \succ x_1 \succ x_4 \succ x_3 \succ x_2$ | 1.101 |

Based on the quantitative analysis results presented in Table 19, our method demonstrates significant advantages in multiple aspects, clearly outperforming other methods.

Firstly, in terms of total execution time, our method requires only 0.0234 s, which is slightly higher than the methods proposed by Liu et al. [55] and Xu et al. [39], but significantly lower than the method by Xu et al. [11] and Liang et al. [54]. This indicates that our method has a high level of computational efficiency, capable of completing tasks in a shorter time. Secondly, regarding the number of iterations, our method only requires 1 iteration to reach the final result, which is the least among all methods. This not only enhances computational efficiency but also reduces the consumption of computational resources, highlighting the advantages of our method in algorithm optimization.

Lastly, in terms of the final ranking result, our method yields the ranking $x_0 \succ x_1 \succ x_4 \succ x_3 \succ x_2$, consistent with the result of Xu et al. [39]'s method and superior to those of Liu et al. [19] and Xu et al. [11]. This demonstrates the high reliability of our method

in terms of decision-making accuracy. In summary, our method excels in execution time, number of iterations, and final ranking results, showcasing high computational efficiency and decision-making accuracy. It is a method worthy of promotion and application.

## 6. Conclusions

This study addresses critical challenges in LSGDM via an enhanced consensus-reaching framework that integrates SNA, refined clustering techniques, a DIS, and minority opinion management. Below are the key contributions and findings:

(1) The proposed method leverages the Hamacher aggregation operator with a decay factor to effectively complete incomplete trust matrices while mitigating biases inherent in traditional statistical imputation. This approach enhances the interpretability and reliability of trust relationships in SNs.
(2) By integrating hybrid distance metrics (Euclidean, Manhattan, and Hamming distances) into the K-means++ algorithm, we achieve fine-grained clustering of DMs. This multidimensional strategy accounts for both evaluation data and preference rankings, significantly improving subgroup homogeneity and reducing computational complexity.
(3) A DIS grounded in MGRSs provides a robust, formula-driven mechanism for assessing consensus levels. By replacing subjective thresholds with granular approximations, this method ensures fairness and transparency in conflict detections.
(4) A systematic mechanism for detecting and adjusting minority opinions is proposed, incorporating subgroup credibility and trust centrality. This balances minority influence without over-amplification, fostering innovation while preserving decision stability.

Empirical validation using UAV procurement datasets demonstrates the framework's effectiveness. Sensitivity analyses confirm the stability of parameter choices, while comparative experiments highlight the superiority of the proposed method over existing approaches in terms of clustering accuracy, consensus efficiency, and ranking rationality.

Potential extensions include adapting the framework to dynamic real-time decision scenarios and optimizing trust propagation via deep learning architectures. Investigating cross-cultural variations in MOH can further enhance global applicability.

By harmonizing technical rigor with practical adaptability, this research advances LSGDM methodology, offering a scalable solution for complex, socially embedded decision challenges. The integration of theoretical innovation and empirical validation establishes a new benchmark in consensus-driven decision systems.

The proposed method offers DMs a more refined tool to reconcile opinion divergences and incorporate dissent constructively—key in high-stakes domains such as public policy, healthcare, or technology procurement. By operationalizing user-generated content through sentiment analysis and trust modeling, the system supports decisions that are both data-driven and socially grounded. The application to UAV selection illustrates how organizations can leverage multi-source evaluations to make informed, balanced choices.

While the proposed consensus-reaching framework demonstrates significant theoretical and empirical advantages, there are several aspects that could be further improved:

(1) The current model is demonstrated in the context of UAV selection based on structured and semi-structured data. Future extensions may consider more heterogeneous decision scenarios, such as new energy vehicles [56].
(2) Although we incorporated social trust propagation, the model does not yet address trust decay over time or dynamic trust formation, which are relevant in evolving SNs.
(3) While our current work integrates sentiment analysis indirectly, future implementations could enhance this component with real-time emotion detection and contextual semantic embeddings.

For further research, one direction is to adapt the framework to dynamic real-time decision scenarios. This could involve developing algorithms that can update trust relationships, clustering results, and consensus levels in real-time as the decision-making environment changes. Another direction is to optimize trust propagation via deep-learning architectures. Deep learning can potentially capture more complex relationships in SNs and decision-making data, improving the accuracy of trust propagation. Additionally, investigating cross-cultural variations in MOH can enhance the global applicability of the model. Different cultures may have different attitudes towards minority opinions, and understanding these differences can help the model better adapt to various cultural backgrounds.

**Author Contributions:** Conceptualization, X.H. and C.Z.; Methodology, C.Z.; Validation, X.H. and T.X.; Data curation, X.H.; Writing—original draft, X.H.; Writing—review & editing, C.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Central Government Guides Local Science and Technology Innovation (YDZJSX2024D015), the Special Fund for the Training Program for Young Scientific Researchers of Higher Education Institutions in Shanxi, the Wenying Young Scholars of Shanxi University, the Science and Technology Innovation Teams of Shanxi (202204051001015), the 22nd Undergraduate Innovation and Entrepreneurship Training Program of Shanxi University (202410108008), and the Cultivate Scientific Research Excellence Programs of Higher Education Institutions in Shanxi (2019SK036).

**Data Availability Statement:** The original contributions presented in this study are included in the article.

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

## Abbreviations

The following abbreviations are used in this paper:

| | |
|---|---|
| LSGDM | Large-scale group decision-making |
| DMs | Decision-makers |
| SNs | Social networks |
| SNA | Social network analysis |
| CRP | Consensus-reaching process |
| MOH | Minority opinion handling |
| DISs | Decision indicator sets |
| IFNs | Intuitionistic fuzzy numbers |
| IFSs | Intuitionistic fuzzy sets |
| SSE | Sum of squared errors |
| UAVs | Unmanned aerial vehicles |
| MGRS | Multigranulation rough set |
| TL | Traditional LSGDM models |
| SNL | Enhanced LSGDM models integrated with SN |
| ML | Behavior-management-oriented LSGDM models to handle misbehaviors |

## Appendix A

**DMs' Evaluation Matrices**

$$\begin{bmatrix} (0.15,0.09) & (0.43,0.12) & (0.13,0.08) & (0.13,0.09) & (0.43,0.12) \\ (0.05,0.04) & (0.40,0.13) & (0.38,0.13) & (0.38,0.13) & (0.26,0.12) \\ (0.61,0.10) & (0.44,0.12) & (0.0,1.0) & (0.26,0.12) & (0.31,0.12) \\ (0.07,0.11) & (0.51,0.12) & (0.60,0.11) & (0.56,0.11) & (1.0,0.0) \\ (0.39,0.13) & (0.71,0.08) & (0.38,0.13) & (0.41,0.12) & (0.38,0.13) \end{bmatrix} \begin{bmatrix} (0.0,0.0) & (0.47,0.12) & (0.81,0.06) & (0.31,0.12) & (0.78,0.07) \\ (0.05,0.07) & (0.41,0.13) & (0.06,0.05) & (0.17,0.10) & (0.18,0.10) \\ (0.04,0.04) & (0.44,0.12) & (0.39,0.13) & (0.0,0.0) & (0.14,0.09) \\ (0.07,0.06) & (0.45,0.12) & (0.16,0.10) & (0.24,0.12) & (0.52,0.12) \\ (0.89,0.04) & (0.47,0.12) & (0.22,0.11) & (0.05,0.04) & (0.17,0.10) \end{bmatrix}$$

$$\begin{bmatrix} (0.40,0.13) & (1.0,0.0) & (0.34,0.12) & (0.08,0.06) & (0.31,0.12) \\ (0.33,0.12) & (0.84,0.05) & (0.47,0.12) & (0.26,0.12) & (0.46,0.12) \\ (0.38,0.13) & (0.44,0.12) & (0.28,0.12) & (0.0,0.0) & (0.80,0.06) \\ (0.82,0.06) & (0.45,0.12) & (0.43,0.12) & (1.0,0.0) & (0.60,0.11) \\ (1.0,0.0) & (0.47,0.12) & (1.0,0.0) & (0.12,0.33) & (0.33,0.12) \end{bmatrix} \begin{bmatrix} (0.92,0.03) & (0.46,0.12) & (0.63,0.10) & (0.56,0.11) & (0.53,0.12) \\ (0.33,0.12) & (0.28,0.12) & (0.43,0.12) & (0.19,0.10) & (0.35,0.13) \\ (0.38,0.13) & (0.38,0.13) & (0.07,0.10) & (0.48,0.12) & (0.12,0.26) \\ (0.40,0.13) & (0.82,0.06) & (0.43,0.12) & (0.54,0.11) & (0.71,0.08) \\ (0.13,0.38) & (0.59,0.11) & (0.18,0.10) & (0.21,0.11) & (0.06,0.08) \end{bmatrix}$$

$$\begin{bmatrix} (0.53,0.12) & (0.46,0.12) & (0.16,0.09) & (0.36,0.13) & (0.52,0.12) \\ (1.0,0.0) & (0.75,0.07) & (0.45,0.12) & (0.57,0.11) & (0.82,0.06) \\ (0.82,0.06) & (0.88,0.04) & (0.62,0.10) & (0.80,0.06) & (0.86,0.04) \\ (0.84,0.05) & (0.82,0.06) & (0.81,0.06) & (0.64,0.10) & (0.78,0.06) \\ (0.05,0.04) & (0.09,0.15) & (0.12,0.31) & (0.12,0.31) & (0.12,0.46) \end{bmatrix} \begin{bmatrix} (0.70,0.09) & (0.83,0.05) & (0.54,0.11) & (0.61,0.10) & (0.66,0.09) \\ (0.77,0.07) & (0.95,0.02) & (0.93,0.02) & (0.90,0.03) & (0.88,0.04) \\ (0.16,0.10) & (0.12,0.33) & (0.08,0.11) & (0.09,0.16) & (0.06,0.05) \\ (0.39,0.13) & (0.22,0.11) & (0.0,0.0) & (0.56,0.11) & (0.25,0.12) \\ (0.79,0.06) & (0.82,0.06) & (0.62,0.10) & (0.29,0.12) & (0.89,0.04) \end{bmatrix}$$

$$\begin{bmatrix} (0.66,0.09) & (0.77,0.07) & (0.80,0.06) & (0.93,0.02) & (0.74,0.07) \\ (0.53,0.12) & (0.59,0.11) & (0.53,0.11) & (0.53,0.12) & (0.57,0.11) \\ (0.38,0.13) & (0.75,0.07) & (0.48,0.12) & (0.20,0.11) & (0.16,0.10) \\ (1.0,0.0) & (1.0,0.0) & (0.70,0.08) & (0.91,0.03) & (0.80,0.06) \\ (0.35,0.13) & (0.06,0.05) & (0.73,0.08) & (0.48,0.12) & (0.30,0.12) \end{bmatrix} \begin{bmatrix} (0.83,0.05) & (0.82,0.06) & (0.86,0.04) & (0.74,0.08) & (0.54,0.11) \\ (1.0,0.0) & (0.97,0.01) & (0.91,0.03) & (1.0,0.0) & (0.97,0.01) \\ (0.85,0.05) & (0.77,0.07) & (0.92,0.03) & (0.03,0.03) & (0.96,0.01) \\ (0.35,0.13) & (0.09,0.07) & (0.22,0.11) & (0.75,0.07) & (1.0,0.0) \\ (0.84,0.05) & (1.0,0.0) & (0.67,0.09) & (0.18,0.10) & (0.69,0.09) \end{bmatrix}$$

$$\begin{bmatrix} (0.28,0.12) & (0.31,0.12) & (0.35,0.13) & (0.27,0.12) & (0.52,0.12) \\ (0.75,0.07) & (1.0,0.0) & (0.54,0.11) & (0.74,0.07) & (1.0,0.0) \\ (0.61,0.10) & (1.0,0.0) & (0.13,0.40) & (0.60,0.11) & (0.76,0.07) \\ (0.61,0.10) & (1.0,0.0) & (0.20,0.11) & (0.16,0.09) & (0.84,0.05) \\ (0.95,0.02) & (0.75,0.07) & (0.34,0.12) & (1.0,0.0) & (0.69,0.09) \end{bmatrix} \begin{bmatrix} (0.54,0.11) & (0.76,0.07) & (0.35,0.13) & (0.28,0.12) & (0.69,0.09) \\ (0.47,0.12) & (0.75,0.07) & (0.46,0.12) & (0.22,0.11) & (0.48,0.12) \\ (0.75,0.07) & (0.44,0.12) & (0.59,0.11) & (0.64,0.10) & (0.78,0.07) \\ (0.52,0.12) & (0.23,0.11) & (0.51,0.12) & (0.46,0.12) & (0.64,0.10) \\ (0.66,0.09) & (0.51,0.12) & (0.76,0.07) & (0.32,0.12) & (0.72,0.08) \end{bmatrix}$$

$$\begin{bmatrix} (0.16,0.10) & (0.76,0.07) & (1.0,0.0) & (0.76,0.07) & (0.26,0.12) \\ (0.22,0.11) & (0.16,0.09) & (0.50,0.12) & (0.03,0.03) & (0.26,0.12) \\ (0.70,0.09) & (0.71,0.08) & (0.19,0.10) & (0.21,0.11) & (0.58,0.11) \\ (0.84,0.05) & (0.54,0.11) & (0.86,0.04) & (0.12,0.32) & (0.80,0.06) \\ (0.21,0.11) & (0.38,0.13) & (0.11,0.08) & (0.22,0.11) & (0.60,0.11) \end{bmatrix} \begin{bmatrix} (0.33,0.12) & (0.16,0.09) & (0.30,0.12) & (0.10,0.07) & (0.39,0.13) \\ (0.62,0.10) & (0.97,0.01) & (0.76,0.07) & (0.79,0.06) & (0.87,0.04) \\ (0.68,0.09) & (0.88,0.04) & (0.69,0.09) & (0.34,0.12) & (0.68,0.09) \\ (0.81,0.06) & (0.44,0.12) & (0.64,0.10) & (0.23,0.11) & (0.56,0.11) \\ (0.49,0.12) & (0.35,0.13) & (0.10,0.07) & (0.21,0.11) & (0.27,0.12) \end{bmatrix}$$

$$\begin{bmatrix} (0.66,0.09) & (0.07,0.05) & (0.56,0.11) & (0.11,0.22) & (0.32,0.12) \\ (0.66,0.09) & (0.18,0.10) & (0.72,0.08) & (0.38,0.13) & (0.88,0.04) \\ (0.02,0.03) & (0.03,0.03) & (0.16,0.09) & (0.05,0.04) & (0.03,0.03) \\ (0.21,0.11) & (0.60,0.11) & (0.09,0.06) & (0.05,0.07) & (0.05,0.04) \\ (0.65,0.09) & (0.69,0.09) & (0.56,0.11) & (0.51,0.12) & (0.88,0.04) \end{bmatrix}$$

## Appendix B

**Trust Degree Matrix Among DMs (Completed)**

$$\begin{pmatrix} 1 & 0.485 & 0.5 & 0.02 & 0.023 & 0.5 & 0.5 & 0.5 & 0.458 & 0.447 & 0.995 & 0.946 & 0.5 \\ 0.66 & 1 & 0.5 & 0.119 & 0.833 & 0.097 & 0.549 & 0.816 & 0.524 & 0.506 & 0.5 & 0.52 & 0.5 \\ 0.5 & 0.115 & 1 & 0.861 & 0.799 & 0.5 & 0.911 & 0.401 & 0.5 & 0.598 & 0.675 & 0.924 & 0.192 \\ 0.13 & 0.473 & 0.968 & 1 & 0.97 & 0.5 & 0.5 & 0.73 & 0.172 & 0.5 & 0.927 & 0.135 & 0.656 \\ 0.26 & 0.251 & 0.677 & 0.975 & 1 & 0.5 & 0.538 & 0.916 & 0.371 & 0.351 & 0.984 & 0.467 & 0.883 \\ 0.228 & 0.788 & 0.5 & 0.817 & 0.282 & 1 & 0.5 & 0.445 & 0.115 & 0.5 & 0.027 & 0.194 & 0.238 \\ 0.317 & 0.152 & 0.387 & 0.5 & 0.808 & 0.168 & 1 & 0.056 & 0.365 & 0.46 & 0.876 & 0.84 & 0.5 \\ 0.201 & 0.5 & 0.598 & 0.039 & 0.63 & 0.5 & 0.772 & 1 & 0.378 & 0.787 & 0.888 & 0.84 & 1 \\ 0.954 & 0.5 & 0.007 & 0.356 & 0.5 & 0.858 & 0.845 & 0.5 & 1 & 0.578 & 0.38 & 0.513 & 0.805 \\ 0.5 & 0.422 & 0.5 & 0.5 & 0.419 & 0.177 & 0.993 & 0.5 & 0.851 & 1 & 0.54 & 0.839 & 0.677 \\ 0.395 & 0.5 & 0.877 & 0.5 & 0.039 & 0.5 & 0.5 & 0.051 & 0.93 & 0.064 & 1 & 0.5 & 0.618 \\ 0.065 & 0.689 & 0.49 & 0.46 & 0.857 & 0.342 & 0.537 & 0.968 & 0.5 & 0.193 & 0.539 & 1 & 0.504 \\ 0.5 & 0.116 & 0.5 & 0.5 & 0.143 & 0.5 & 0.76 & 0.5 & 0.5 & 0.197 & 0.066 & 0.561 & 1 \end{pmatrix}$$

## Appendix C

**Sub-Group Evaluation Matrices**

$$\begin{bmatrix} (0.22,0.06) & (0.28,0.04) & (0.19,0.05) & (0.17,0.04) & (0.26,0.05) \\ (0.30,0.04) & (0.39,0.03) & (0.26,0.05) & (0.26,0.05) & (0.34,0.04) \\ (0.29,0.05) & (0.36,0.03) & (0.23,0.07) & (0.20,0.04) & (0.32,0.04) \\ (0.38,0.03) & (0.32,0.03) & (0.27,0.05) & (0.28,0.04) & (0.34,0.04) \\ (0.28,0.03) & (0.17,0.05) & (0.25,0.05) & (0.18,0.08) & (0.19,0.08) \end{bmatrix} \quad \begin{bmatrix} (0.83,0.05) & (0.82,0.06) & (0.86,0.04) & (0.74,0.08) & (0.54,0.11) \\ (1.00,0.00) & (0.97,0.01) & (0.91,0.03) & (1.00,0.00) & (0.97,0.01) \\ (0.85,0.05) & (0.77,0.07) & (0.92,0.03) & (0.03,0.03) & (0.96,0.01) \\ (0.35,0.13) & (0.09,0.07) & (0.22,0.11) & (0.75,0.07) & (1.00,0.00) \\ (0.84,0.05) & (1.00,0.00) & (0.67,0.09) & (0.18,0.10) & (0.69,0.09) \end{bmatrix}$$

$$\begin{bmatrix} (0.10,0.01) & (0.06,0.01) & (0.08,0.02) & (0.05,0.02) & (0.07,0.02) \\ (0.10,0.01) & (0.08,0.01) & (0.12,0.01) & (0.09,0.01) & (0.13,0.01) \\ (0.01,0.01) & (0.01,0.02) & (0.02,0.02) & (0.01,0.01) & (0.01,0.01) \\ (0.04,0.02) & (0.06,0.02) & (0.01,0.01) & (0.04,0.01) & (0.02,0.01) \\ (0.10,0.01) & (0.11,0.01) & (0.09,0.02) & (0.06,0.02) & (0.13,0.01) \end{bmatrix} \quad \begin{bmatrix} (0.09,0.02) & (0.16,0.03) & (0.20,0.02) & (0.14,0.03) & (0.14,0.03) \\ (0.05,0.03) & (0.09,0.03) & (0.10,0.03) & (0.05,0.02) & (0.08,0.03) \\ (0.13,0.03) & (0.15,0.03) & (0.05,0.09) & (0.07,0.03) & (0.09,0.04) \\ (0.11,0.02) & (0.17,0.03) & (0.16,0.03) & (0.10,0.05) & (0.22,0.02) \\ (0.11,0.05) & (0.15,0.03) & (0.06,0.03) & (0.06,0.03) & (0.09,0.03) \end{bmatrix}$$

## References


1. Kiesler, S.; Sproull, L. Group decision making and communication technology. *Organ. Behav. Hum. Decis. Process.* **1992**, *52*, 96–123. [CrossRef]
2. Wang, Y.; Song, H.; Dutta, B.; García-Zamora, D.; Martínez, L. Consensus reaching in LSGDM: Overlapping community detection and bounded confidence-driven feedback mechanism. *Inf. Sci.* **2024**, *679*, 121104. [CrossRef]
3. Tang, M.; Liao, H. From conventional group decision making to large-scale group decision making: What are the challenges and how to meet them in big data era? A state-of-the-art survey. *Omega* **2021**, *100*, 102141. [CrossRef]
4. Zhang, C.; Wang, Y.; Sangaiah, A.; Alenazi, M.J.F.; Aborokbah, M. An incomplete three-way consensus algorithm for unmanned aerial vehicle purchase using optimization-driven sentiment analysis. *Future Gener. Comput. Syst.* **2025**, 107761. [CrossRef]
5. Garcia-Zamora, D.; Labella, A.; Ding, W.; Rodríguez, R.M.; Martínez, L. Large-scale group decision making: A systematic review and a critical analysis. *IEEE/CAA J. Autom. Sin.* **2022**, *9*, 949–966. [CrossRef]
6. Li, X. Big data-driven fuzzy large-scale group decision making (LSGDM) in circular economy environment. *Technol. Forecast. Soc. Change* **2022**, *175*, 121285.
7. Shen, Y.; Ma, X.; Xu, Z.; Herrera-Viedma, E.; Maresova, P.; Zhan, J. Opinion evolution and dynamic trust-driven consensus model in large-scale group decision-making under incomplete information. *Inf. Sci.* **2024**, *657*, 119925. [CrossRef]
8. Mahmoudi, A.; Abbasi, M.; Yuan, J.; Li, L. Large-scale group decision-making (LSGDM) for performance measurement of healthcare construction projects: Ordinal Priority Approach. *Appl. Intell.* **2022**, *52*, 13781–13802. [CrossRef]
9. Zhang, R.; Huang, J.; Xu, Y. Consensus models with aggregation operators for minimum quadratic cost in group decision making. *Appl. Intell.* **2023**, *53*, 1370–1390. [CrossRef]
10. Wang, Y.; He, S.; Zamora, D.; Pan, X.; Martínez, L. A large scale group three-way decision-based consensus model for site selection of new energy vehicle charging stations. *Expert Syst. Appl.* **2023**, *214*, 119107. [CrossRef]
11. Xu, T.; He, S.; Yuan, X.; Zhang, C. Enhancing group consensus in social networks: A two-stage dual-fine tuning consensus model based on adaptive Leiden algorithm and minority opinion management with non-cooperative behaviors. *Electronics* **2024**, *13*, 4930. [CrossRef]
12. Hsieh, Y.; Lu, L.; Ku, Y. Review evaluation for hotel recommendation. *Electronics* **2023**, *12*, 4673. [CrossRef]
13. Tian, X.; Ma, W.; Wu, L.; Xie, M.; Kou, G. Large-scale consensus with dynamic trust and optimal reference in social network under incomplete probabilistic linguistic circumstance. *Inf. Sci.* **2024**, *661*, 120123. [CrossRef]
14. Harshavardhan, A.; Boyapati, P.; Neelakandan, S.; Abdul-Rasheed Akeji, A.A.; Singh Pundir, A.K.; Walia, R. LSGDM with Biogeography-Based Optimization (BBO) Model for Healthcare Applications. *J. Healthc. Eng.* **2022**, *2022*, 2170839. [CrossRef]
15. Chunaev, P. Community detection in node-attributed social networks: A survey. *Comput. Sci. Rev.* **2020**, *37*, 100286. [CrossRef]
16. Jin, F.; Cai, Y.; Zhou, L.; Ding, T. Regret–rejoice two-stage multiplicative DEA models-driven cross-efficiency evaluation with probabilistic linguistic information. *Omega* **2023**, *117*, 102839. [CrossRef]
17. Wu, J.; Chiclana, F.; Fujita, H.; Herrera-Viedma, E. A visual interaction consensus model for social network group decision making with trust propagation. *Knowl.-Based Syst.* **2017**, *122*, 39–50. [CrossRef]
18. Liu, B.; Zhou, Q.; Ding, R.X.; Palomares, I.; Herrera, F. Large-scale group decision making model based on social network analysis: Trust relationship-based conflict detection and elimination. *Eur. J. Oper. Res.* **2019**, *275*, 737–754. [CrossRef]
19. Liu, P.; Li, Y.; Wang, P. Social trust-driven consensus reaching model for multiattribute group decision making: Exploring social trust network completeness. *IEEE Trans. Fuzzy Syst.* **2023**, *31*, 3040–3054. [CrossRef]
20. Chu, J.; Wang, Y.; Liu, X.; Liu, Y. Social network community analysis based large-scale group decision making approach with incomplete fuzzy preference relations. *Inf. Fusion* **2020**, *60*, 98–120. [CrossRef]
21. Xu, Y.; Li, C.; Wen, X. Missing values estimation and consensus building for incomplete hesitant fuzzy preference relations with multiplicative consistency. *Int. J. Comput. Intell. Syst.* **2018**, *11*, 101–119. [CrossRef]

22. Song, Y.; Li, G. Handling group decision-making model with incomplete hesitant fuzzy preference relations and its application in medical decision. *Soft Comput.* **2019**, *23*, 6657–6666. [CrossRef]
23. Liu, P. Some Hamacher aggregation operators based on the interval-valued intuitionistic fuzzy numbers and their application to group decision making. *IEEE Trans. Fuzzy Syst.* **2013**, *22*, 83–97. [CrossRef]
24. Ren, Y.; Pu, J.; Yang, Z.; Xu, J.; Li, G.; Pu, X.; Yu, P.S.; He, L. Deep clustering: A comprehensive survey. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *36*, 5858–5878. [CrossRef]
25. Xu, D.; Tian, Y. A comprehensive survey of clustering algorithms. *Ann. Data Sci.* **2015**, *2*, 165–193. [CrossRef]
26. Wang, Y.; Qian, J.; Hassan, M.; Zhang, X.; Zhang, T.; Yang, C.; Zhou, X.; Jia, F. Density peak clustering algorithms: A review on the decade 2014–2023. *Expert Syst. Appl.* **2024**, *238*, 121860. [CrossRef]
27. Kapoor, A.; Singhal, A. A comparative study of K-Means, K-Means++ and Fuzzy C-Means clustering algorithms. In Proceedings of the 2017 3rd International Conference on Computational Intelligence & Communication Technology (CICT), Ghaziabad, India, 9–10 February 2017; pp. 1–6.
28. Hajihosseinlou, M.; Maghsoudi, A.; Ghezelbash, R. Intelligent mapping of geochemical anomalies: Adaptation of DBSCAN and mean-shift clustering approaches. *J. Geochem. Explor.* **2024**, *258*, 107393. [CrossRef]
29. Cheng, F. A comparative study of the performance of Spark-based k-means algorithm based on Euclidean distance and Manhattan distance. In Proceedings of the 2024 3rd International Conference on Big Data, Information and Computer Network (BDICN), Sanya, China, 12–14 January 2024; pp. 1–6.
30. Guo, L.; Zhan, J.; Zhang, C.; Xu, Z. A large-scale group decision-making method fusing three-way clustering and regret theory under fuzzy preference relations. *IEEE Trans. Fuzzy Syst.* **2023**, *32*, 4846–4860. [CrossRef]
31. Labella, A.; Liu, H.; Rodríguez, R.M.; Martínez, L. A cost consensus metric for consensus reaching processes based on a comprehensive minimum cost model. *Eur. J. Oper. Res.* **2020**, *281*, 316–331. [CrossRef]
32. Zou, W.; Wan, S.; Dong, J.; Martínez, L. A new social network driven consensus reaching process for multi-criteria group decision making with probabilistic linguistic information. *Inf. Sci.* **2023**, *632*, 467–502. [CrossRef]
33. Tang, M.; Liao, H.; Mi, X.; Lev, B.; Pedrycz, W. A hierarchical consensus reaching process for group decision making with noncooperative behaviors. *Eur. J. Oper. Res.* **2021**, *293*, 632–642. [CrossRef]
34. Cheng, X.; Zhang, K.; Wu, T.; Xu, Z.; Gou, X. An opinions-updating model for large-scale group decision-making driven by autonomous learning. *Inf. Sci.* **2024**, *662*, 120238. [CrossRef]
35. Gai, T.; Wu, J.; Cao, M.; Liu, Y.; Liang, C. Blockchain platform selection for supply chain finance: A bilateral-negotiation-based group multiattribute decision making method. *IEEE Trans. Comput. Soc. Syst.* **2024**, *11*, 6072–6086. [CrossRef]
36. Zhang, X.; Liu, Q.; Zhang, C.; Zhan, J. New fuzzy rough set models based on implication operators. *Comput. Appl. Math.* **2025**, *44*, 125. [CrossRef]
37. Laor, T. Breaking the silence: The role of social media in fostering community and challenging the spiral of silence. *Online Inf. Rev.* **2024**, *48*, 710–724. [CrossRef]
38. Ren, R.; Tang, M.; Liao, H. Managing minority opinions in micro-grid planning by a social network analysis-based large scale group decision making method with hesitant fuzzy linguistic information. *Knowl.-Based Syst.* **2020**, *189*, 105060. [CrossRef]
39. Xu, X.; Du, Z.; Chen, X. Consensus model for multi-criteria large-group emergency decision making considering non-cooperative behaviors and minority opinions. *Decis. Support Syst.* **2015**, *79*, 150–160. [CrossRef]
40. Liu, P.; Li, Y.; Wang, P. Opinion dynamics and minimum adjustment-driven consensus model for multi-criteria large-scale group decision making under a novel social trust propagation mechanism. *IEEE Trans. Fuzzy Syst.* **2022**, *31*, 307–321. [CrossRef]
41. Chaira, T. A novel intuitionistic fuzzy C means clustering algorithm and its application to medical images. *Appl. Soft Comput.* **2011**, *11*, 1711–1717. [CrossRef]
42. Shen, Y.; Ma, X.; Zhang, H.; Zhan, J. Fusion social network and regret theory for a consensus model with minority opinions in large-scale group decision making. *Inf. Fusion* **2024**, *112*, 102548. [CrossRef]
43. Zhang, Z.; Xu, Y.; Hao, W.; Yu, X. A signed network analysis-based consensus reaching process in group decision making. *Appl. Soft Comput.* **2021**, *100*, 106926. [CrossRef]
44. Zhang, H.; Palomares, I.; Dong, Y.; Wang, W. Managing non-cooperative behaviors in consensus-based multiple attribute group decision making: An approach based on social network analysis. *Knowl.-Based Syst.* **2018**, *162*, 29–45. [CrossRef]
45. Tan, C.; Yi, W.; Chen, X. Hesitant fuzzy Hamacher aggregation operators for multicriteria decision making. *Appl. Soft Comput.* **2015**, *26*, 325–349. [CrossRef]
46. Liu, X.; Yang, Y.; Jiang, J. The behavioral TOPSIS based on prospect theory and regret theory. *Int. J. Inf. Technol. Decis. Mak.* **2023**, *22*, 1591–1615. [CrossRef]
47. Bisht, G.; Pal, A.K. Prospect–regret theory based decision-making approach for incomplete probabilistic hesitant fuzzy environment: An application to medical field. *Expert Syst. Appl.* **2024**, *250*, 123906. [CrossRef]
48. Kahneman, D.; Tversky, A. Prospect theory: An analysis of decision under risk. In *Handbook of the Fundamentals of Financial Decision Making: Part I*; World Scientific: Hackensack, NJ, USA, 2013; pp. 99–127.

49. Zhang, S.; Zhu, J.; Liu, X.; Chen, Y. Regret theory-based group decision-making with multidimensional preference and incomplete weight information. *Inf. Fusion* **2016**, *31*, 1–13. [CrossRef]
50. Li, R.; Zhang, C.; Li, D.; Li, W.; Zhan, J. Improved evidential three-way decisions in incomplete multi-scale information systems. *Int. J. Approx. Reason.* **2025**, *181*, 109417. [CrossRef]
51. Yuan, X.; Xu, T.; He, S.; Zhang, C. An online review data-driven fuzzy large-scale group decision-making method based on dual fine-tuning. *Electronics* **2024**, *13*, 2702. [CrossRef]
52. Chen, G.; Wei, L.; Fu, J.; Li, C.; Zhao, G. A large group emergency decision-making method based on uncertain linguistic cloud similarity method. *Math. Comput. Appl.* **2022**, *27*, 101. [CrossRef]
53. Meng, F.; Tang, J.; An, Q. Cooperative game based two-stage consensus adjustment mechanism for large-scale group decision making. *Omega* **2023**, *117*, 102842. [CrossRef]
54. Liang, X.; Guo, J.; Liu, P. A large-scale group decision-making model with no consensus threshold based on social network analysis. *Inf. Sci.* **2022**, *612*, 361–383. [CrossRef]
55. Liu, Z.; Wang, W.; Zhang, X.; Liu, P. A dynamic dual-trust network-based consensus model for individual non-cooperative behaviour management in group decision-making. *Inf. Sci.* **2024**, *674*, 120750. [CrossRef]
56. Ding, J.; Zhang, C.; Li, D.; Li, W.; Zhan, J. A three-way large-scale group decision-making method integrating sentiment analysis and quantum interference-based prospect theory for the selection of new energy vehicles. *Expert Syst. Appl.* **2025**, *275*, 126940. [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.

*electronics*

MDPI

*Review*

# Integrating Artificial Intelligence Agents with the Internet of Things for Enhanced Environmental Monitoring: Applications in Water Quality and Climate Data

**Tymoteusz Miller [1,2,*], Irmina Durlik [3], Ewelina Kostecka [4], Polina Kozlovska [5], Adrianna Łobodzińska [6,7], Sylwia Sokołowska [8] and Agnieszka Nowy [3]**

1 Institute of Marine and Environmental Sciences, University of Szczecin, 71-141 Szczecin, Poland
2 Faculty of Data Science and Information, INTI International University, Nilai 71800, Negeri Sembilan, Malaysia
3 Faculty of Navigation, Maritime University of Szczecin, 71-650 Szczecin, Poland
4 Faculty of Mechatronics and Electrical Engineering, Maritime University of Szczecin, 71-650 Szczecin, Poland
5 Faculty of Economics, Finance and Management, University of Szczecin, 71-101 Szczecin, Poland
6 Institute of Biology, University of Szczecin, 71-415 Szczecin, Poland
7 Doctoral School of the University of Szczecin, 71-412 Szczecin, Poland
8 Polish Society of Bioinformatics and Data Science BioData, 71-214 Szczecin, Poland
* Correspondence: tymoteusz.miller@usz.edu.pl

**Abstract:** The integration of artificial intelligence (AI) agents with the Internet of Things (IoT) has marked a transformative shift in environmental monitoring and management, enabling advanced data gathering, in-depth analysis, and more effective decision making. This comprehensive literature review explores the integration of AI and IoT technologies within environmental sciences, with a particular focus on applications related to water quality and climate data. The methodology involves a systematic search and selection of relevant studies, followed by thematic, meta-, and comparative analyses to synthesize current research trends, benefits, challenges, and gaps. The review highlights how AI enhances IoT's data collection capabilities through advanced predictive modeling, real-time analytics, and automated decision making, thereby improving the accuracy, timeliness, and efficiency of environmental monitoring systems. Key benefits identified include enhanced data precision, cost efficiency, scalability, and the facilitation of proactive environmental management. Nevertheless, this integration encounters substantial obstacles, including issues related to data quality, interoperability, security, technical constraints, and ethical concerns. Future developments point toward enhancements in AI and IoT technologies, the incorporation of innovations like blockchain and edge computing, the potential formation of global environmental monitoring systems, and greater public involvement through citizen science initiatives. Overcoming these challenges and embracing new technological trends could enable AI and IoT to play a pivotal role in strengthening environmental sustainability and resilience.

**Keywords:** artificial intelligence (AI) agents; Internet of Things (IoT); environmental monitoring; water quality; climate data; predictive analytics; real-time decision making; sustainable environmental management; AI–IoT integration; smart environmental systems

## 1. Introduction

### *1.1. Background*

The modern world is confronted with numerous environmental issues [1–3], which present serious risks to ecosystems [4,5], public health [6,7], and the global economy [8–10].

Issues such as climate change [11–13], water pollution [14–16], deforestation [17–19], and loss of biodiversity [20,21] are escalating at an unprecedented rate, driven by industrialization [22–24], urbanization [25–27], and unsustainable resource exploitation [28,29]. For instance, rising global temperatures [30–33] are altering weather patterns, leading to extreme events like hurricanes, droughts, and floods, which in turn disrupt agricultural productivity and compromise water security. Likewise, the pollution of water bodies [34,35] due to agricultural runoff, industrial discharges, and untreated sewage not only harms aquatic ecosystems but also presents significant health hazards to communities depending on these water sources.

In addressing these complex and interrelated challenges, timely and accurate decision-making in environmental sciences is paramount. Successful management and mitigation efforts rely on real-time environmental monitoring, accurate trend prediction, and proactive responses to emerging threats. Conventional monitoring methods, which typically involve intermittent data collection and manual analysis, are becoming insufficient given the accelerating pace of environmental changes and the vast amounts of data produced. As a result, there is an urgent demand for innovative technological solutions that can improve the accuracy, efficiency, and scalability of environmental data acquisition and analysis.

### *1.2. Role of Technology in Environmental Monitoring*

The development of environmental monitoring tools [36–38] has significantly progressed in recent decades, fueled by technological advancements and the growing need for comprehensive, real-time environmental data. Early environmental monitoring efforts were largely manual and labor-intensive, involving field surveys, laboratory analyses, and paper-based record-keeping [39,40]. Although these approaches offered important insights, their effectiveness was constrained by limited scope, infrequent data collection, and the risk of human error. The advent of digital technologies has revolutionized environmental monitoring [41,42] by enabling automated data collection, storage, and analysis. The Internet of Things (IoT) has become a key technology in this evolution, enabling the use of interconnected sensors and devices for the continuous monitoring of diverse environmental parameters [43–45]. IoT-based sensors can monitor various environmental parameters, such as air and water quality, temperature, humidity, and soil moisture, while transmitting real-time data to centralized platforms for analysis [46–48]. In conjunction with IoT, Artificial Intelligence (AI) technologies have brought advanced capabilities for processing and interpreting the large datasets produced by environmental sensors [49,50]. Machine learning algorithms can detect patterns, forecast trends, and extract actionable insights from complex environmental data, thereby improving the decision-making process [51–53]. AI-driven models can forecast pollution levels, predict climate anomalies, and optimize resource management strategies, enabling environmental scientists and policymakers to respond swiftly and effectively to emerging challenges. The combination of IoT and AI [54] creates a synergistic framework that harnesses the advantages of both technologies. IoT establishes the foundation for continuous and extensive data collection, while AI delivers the analytical capabilities needed to convert raw data into actionable insights. This integration enables quicker and more precise decision making, ultimately enhancing environmental management and sustainability efforts.

This article aims to examine the integration of artificial intelligence (AI) agents and the Internet of Things (IoT) in environmental sciences, with a particular emphasis on improving decision-making processes concerning water quality and climate data. By examining the interplay between these technologies, this article aims to elucidate how their combined application can address the pressing environmental challenges of our time. Specifically, the article seeks to achieve the following objectives:

1. Define and characterize AI agents within the context of environmental sciences, highlighting their various types, functions, and capabilities.
2. Give a summary of IoT technologies applicable to environmental monitoring, highlighting the various sensors utilized for collecting water quality and climate data.
3. Examine the interaction between IoT and AI, explaining how their integration enhances data accuracy, enables real-time monitoring, and supports proactive decision making.
4. Examine case studies and practical applications where AI and IoT have been successfully implemented to monitor and manage environmental parameters.
5. Explore the advantages, obstacles, and future prospects of implementing integrated AI–IoT systems in environmental sciences.

## 2. Methodology of Literature Review

A systematic literature review was carried out to examine the integration of AI and IoT in environmental sciences, with a focus on water quality and climate monitoring. The review highlights major research trends, effective applications, existing challenges, and potential future developments.

A systematic search was performed across IEEE Xplore, SpringerLink, ScienceDirect, Wiley Online Library, and PubMed, supplemented by industry reports (WHO, EPA, IEA) and conference proceedings. Keywords included "AI agents", "IoT", "environmental monitoring", and "predictive analytics", with Boolean operators to refine results.

Studies published between 2010 and 2024 were included if they addressed AI–IoT integration in environmental sciences, prioritizing peer-reviewed research. Non-scholarly sources, outdated publications, and theoretical-only studies without empirical validation were excluded. Global research diversity was considered to ensure a broad perspective.

Titles and abstracts were screened based on selection criteria, followed by full-text analysis. Extracted data included study objectives, methodologies, AI–IoT technologies used, key findings, and challenges. Thematic analysis categorized studies by AI techniques, IoT applications, benefits, and limitations, while comparative and meta-analysis highlighted trends and best practices.

The synthesis revealed improvements in real-time monitoring, predictive analytics, and automated decision–making through AI–IoT integration. However, challenges such as data quality, interoperability, ethical concerns, and long-term sustainability remain. The review acknowledges limitations, including potential publication bias, language constraints, and the exclusion of purely theoretical advancements. Future research should address scalability, security, interdisciplinary collaboration, and emerging AI techniques like reinforcement learning and federated learning.

## 3. Understanding AI Agents in Environmental Sciences

Artificial intelligence (AI) agents have become essential instruments in advancing environmental sciences [55–57]. Leveraging their diverse capabilities, these agents improve monitoring [58–60], facilitate analysis [61–64], and optimize decision-making processes [55,65,66], all of which are crucial for tackling complex environmental issues. This chapter provides an in-depth examination of AI agents, outlining their definitions, classifications, functionalities, and practical applications in environmental settings (Figure 1).

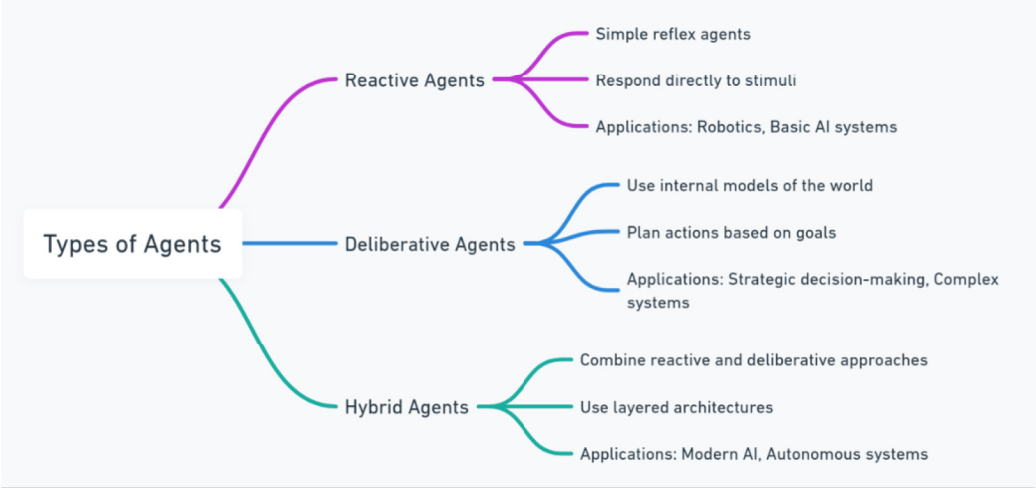



**Figure 1.** Classification of AI agents in environmental sciences.

### *3.1. Definition and Types of AI Agents*

AI agents are self-governing systems capable of sensing their surroundings, processing data, and executing actions to accomplish predefined goals. In environmental sciences, they support the handling and interpretation of large datasets, allowing for more efficient and timely decision making. Based on their operational mechanisms and decision-making approaches, AI agents are generally classified into three main types: reactive agents, deliberative agents, and hybrid agents [67]. Reactive agents [68,69] operate by responding directly to environmental stimuli based on predefined rules. They lack the capacity to maintain internal states or plan for future actions, making their behavior entirely dependent on the current input from their sensors. In environmental applications, reactive agents are commonly employed for real-time monitoring and alerting systems [70]. For example, a reactive AI agent might activate an alarm when sensor data indicate a sudden increase in water pollution levels, facilitating immediate remedial actions [71]. These agents are valued for their simplicity and speed, although they are limited in their adaptability and lack of long-term planning capabilities. Deliberative agents [72], on the other hand, possess the ability to plan and reason about future actions based on internal models of their environment. Unlike reactive agents, deliberative agents maintain an internal state that allows them to consider the consequences of their actions over time [73]. In environmental sciences, deliberative agents are utilized for tasks requiring strategic planning and long-term decision making, such as climate modeling and ecosystem management [74]. For example, a deliberative AI agent can assess historical climate data to forecast future temperature patterns and recommend appropriate mitigation strategies [75]. These agents are characterized by their complexity and adaptability, though they require more computational resources and extensive data for effective operation [76]. Hybrid agents [77,78] combine the strengths of both reactive and deliberative agents, enabling them to respond swiftly to immediate changes while also engaging in long-term planning. This dual capability makes hybrid agents particularly suitable for dynamic and complex environmental systems [79] where both real-time responsiveness and strategic foresight are necessary. For example, a hybrid AI agent deployed in a smart irrigation system can adjust water distribution in real-time based on current soil moisture levels while also planning future irrigation [80] schedules based on weather forecasts and crop growth models [81,82]. Hybrid agents offer versatility and enhanced performance by integrating multiple operational modes, thereby providing a superior functionality in various environmental applications (Table 1).

**Table 1.** Comparison of AI agent types.

| AI Agent Type | Characteristics | Functionalities | Advantages | Typical Applications |
|---|---|---|---|---|
| Reactive | Operate based on predefined rules; no internal state or future planning | Real-time monitoring and alerting [70] | Simplicity, speed | Pollution detection [83,84], immediate remedial actions |
| Deliberative | Maintain internal state; capable of planning and reasoning about future actions | Climate modeling, ecosystem management | Complex reasoning, adaptability | Climate forecasting, resource management [85] |
| Hybrid | Combine reactive and deliberative capabilities; handle both immediate and long-term tasks | Smart irrigation systems, autonomous vehicles | Versatility, enhanced performance | Dynamic resource allocation, adaptive systems [40,86] |

### *3.2. Functions and Capabilities*

In environmental sciences, AI agents fulfill various functions that improve the efficiency, precision, and effectiveness of monitoring and management processes [87]. Their capabilities encompass data collection and processing, predictive analytics, and automated decision making, each contributing to a comprehensive approach to environmental stewardship [88]. Data collection and processing represent the core functions of AI agents. By utilizing IoT devices and sensor networks, these agents capture real-time data on key environmental parameters, including air and water quality, temperature, humidity, and pollutant concentrations [89–92]. The ability to process large volumes of data rapidly allows AI agents to filter, clean, and organize information for further analysis. Sensor [93] integration ensures seamless data acquisition from diverse sources, while automated data cleaning identifies and corrects anomalies and inconsistencies in the raw data, maintaining data quality and reliability [91,94,95]. This high-speed data processing capability ensures that environmental scientists have access to up-to-date information essential for accurate assessments and timely interventions. Predictive analytics represents another significant capability of AI agents. By analyzing historical and current data, AI agents can forecast future environmental conditions and trends. This predictive power is crucial for anticipating events such as natural disasters, pollution spikes, and climate anomalies, enabling proactive measures to mitigate adverse impacts. Machine learning models, such as regression, classification, and neural networks, are utilized to detect patterns and generate predictions from the collected data. Scenario simulation allows AI agents to create various environmental scenarios, assessing potential outcomes and informing decision making. Continuous trend analysis ensures that emerging patterns and shifts in environmental parameters are detected promptly, supporting long-term planning and sustainability efforts. Automated decision making strengthens the role of AI agents in environmental sciences by delivering data-driven insights and recommendations. AI algorithms assess available options and identify the most effective actions based on predefined criteria and real-time data inputs. This functionality is especially crucial in situations demanding swift responses, such as emergency management and resource distribution. Decision support systems powered by AI present relevant data and analytical outcomes through visualization tools, aiding human decision makers in interpreting complex information. Optimization algorithms identify the best possible solutions for resource management, pollution control, and conservation

efforts, while autonomous actions implement predefined measures without the need for human intervention, especially in time-sensitive situations.

Algorithm Selection Criteria

The suitability of AI algorithms for environmental monitoring was determined based on three key criteria: data type, computational efficiency, and interpretability.

Data type played a crucial role in algorithm selection. Supervised learning models, such as random forests and support vector machines (SVMs), were used for labeled datasets, such as water quality measurements with known contamination levels. In contrast, unsupervised clustering techniques like K-Means were applied for anomaly detection in unlabeled climate datasets, identifying deviations in air pollution or temperature patterns. Deep learning architectures, such as convolutional neural networks (CNNs), were selected for high-dimensional data, particularly in remote sensing applications where feature extraction from satellite imagery was required.

Computational efficiency was another critical factor, especially for IoT-based monitoring solutions deployed in resource-constrained environments. Random forests and decision trees were preferred for edge devices due to their low computational overhead and real-time processing capabilities, making them ideal for on-device water quality assessments. In contrast, LSTM networks were chosen for time-series water quality prediction, leveraging their ability to capture long-term dependencies in sensor readings while running on cloud-based platforms. CNNs, while computationally intensive, were prioritized for high-resolution environmental image processing, where deep feature extraction provided significant advantages.

Interpretability was essential for ensuring stakeholder trust and regulatory compliance in AI-driven decision making. SHAP (Shapley additive explanations) values were used to explain feature contributions in pollution source identification, providing transparency for environmental policy enforcement. In regulatory contexts, where decision-making accountability was necessary, logistic regression and random forests were preferred due to their explainability, whereas deep learning models were primarily used in cases where accuracy outweighed interpretability concerns.

For example, LSTM networks were prioritized for time-series water quality prediction, effectively modeling seasonal variations and pollution trends in lakes and rivers [96–103]. Meanwhile, K-means clustering was applied to identify spatial pollution hotspots by grouping IoT sensor data based on air quality measurements. These algorithmic choices ensured that AI-driven environmental monitoring systems were optimized for accuracy, efficiency, and real-world applicability.

### *3.3. Real-World Applications of AI Agents in Environmental Monitoring and Management*

AI agents have been utilized across diverse environmental applications, showcasing their potential to enhance sustainability efforts. Their use extends to resource management, biodiversity conservation, urban planning [96], emissions forecasting [97], and advanced robotics, highlighting the adaptability and influence of AI in environmental sciences. In the field of resource management and climate change mitigation [98–100], AI technologies play a crucial role in optimizing resource use, forecasting environmental trends [101,102], and enhancing decision-making processes. For instance, in the Great Lakes region, an AI-powered water quality monitoring system [103] utilizes IoT sensors to continuously measure parameters such as pH, turbidity, dissolved oxygen, and contaminant levels. AI agents process this data in real-time, identifying pollution events and forecasting future water quality trends. This approach has greatly enhanced proactive water quality management, helping to mitigate pollution-related challenges more effectively. Biodiver-

sity preservation has also benefited from AI integration [104,105]. AI-driven drones and camera traps are utilized for wildlife monitoring, detecting potential threats and aiding conservation efforts [106]. Machine learning models evaluate ecosystem health, supporting sustainable management strategies to safeguard biodiversity [107]. However, the implementation of AI in conservation raises ethical and cybersecurity concerns that must be addressed to ensure responsible use and protect sensitive environmental data. Urban planning and green building technologies have seen significant advancements through AI applications [108]. In urban environments, AI algorithms analyze data on population density, land use, and infrastructure development to guide sustainable urban expansion. Machine learning models optimize building designs for energy efficiency, reducing resource consumption and lowering carbon emissions [109]. Additionally, AI systems monitor and control indoor environments, ensuring optimal air quality, temperature, and lighting for occupants. These AI-powered technologies support the creation of smart cities that integrate sustainable development with urban growth, improving residents' quality of life while reducing the environmental impact [110,111]. AI has also been pivotal in predicting $CO_2$ emissions [112,113]. The biogeography-based optimization (BBO) algorithm has been utilized to model and forecast emission trends, providing policymakers with data-driven insights necessary for formulating effective environmental policies and strategies [114]. This strategy enables the implementation of targeted initiatives to lower emissions and counteract climate change effects, ensuring that environmental management practices remain both efficient and sustainable [115,116]. Innovative applications of AI extend to the realm of robotics [117,118], where AI agents equip micro and nanoscale colloidal robots with deep reinforcement learning capabilities. These robots are capable of navigating unknown environments efficiently, making them particularly impactful in precision surgery and targeted nanodrug delivery [119]. AI-driven navigation enables robots to avoid obstacles and minimize travel time based on local sensory inputs, enhancing the precision and effectiveness of medical interventions [120]. Research on autonomous AI agents [121] in open-world environments is another frontier in environmental applications. These agents utilize integrative frameworks that combine reinforcement learning with symbolic planning to handle complex tasks and adapt to new conditions. Such capabilities are crucial for applications like autonomous vehicles and service robots, which operate in unpredictable settings and require robust adaptability to function effectively [122,123]. These autonomous systems represent the cutting edge of AI integration in dynamic and real-world environments, offering solutions that are both intelligent and resilient [124,125]. Finally, AI agents are being incorporated into extended reality (XR) applications [126] to deliver highly detailed and immersive training experiences [127]. In contexts such as LEGO brick assembly, AI agents use large language models and vision–language integration to decide actions based on past experiences [128], enhancing user interaction in XR environments [129]. This integration facilitates more effective and immersive training experiences, improving learning outcomes and user engagement through personalized and responsive AI-driven guidance. AI agents, through their diverse types and sophisticated capabilities, are transforming environmental sciences by enabling more effective monitoring, analysis, and decision making. Reactive, deliberative, and hybrid agents each offer unique advantages tailored to specific environmental applications, from real-time pollution detection to strategic climate modeling. The presented case studies demonstrate the concrete advantages of combining AI agents with IoT technologies, showcasing advancements in water quality monitoring, biodiversity conservation, urban planning, emissions forecasting, and cutting-edge robotics. As AI and IoT continue to progress, their integration is expected to further strengthen the ability of environmental scientists and policymakers to tackle

urgent environmental issues, promoting sustainability and resilience through intelligent and forward-thinking management approaches.

### *3.4. Critical Analysis of Selected Case Studies*

The application of AI in biodiversity conservation, urban planning, and $CO_2$ emissions forecasting has demonstrated significant potential, but critical challenges remain. This section provides an in-depth evaluation of selected case studies, analyzing their strengths, limitations, and areas for improvement. The comparative table below summarizes the key findings, offering a structured perspective on the effectiveness and scalability of AI-driven environmental monitoring solutions Table 2.

**Table 2.** Comparative analysis of AI applications in environmental monitoring.

| Study | Domain | AI Technique | Strengths | Weaknesses |
|---|---|---|---|---|
| Raihan (2023) [104] | Biodiversity conservation | Machine learning | Comprehensive overview of AI in conservation; practical applications for resource management | Lacks empirical validation; limited discussion on AI adoption challenges |
| Ullah, Saqib and Xiong (2024) [105] | Biodiversity conservation | AI for ecosystem monitoring | Bridges classical and modern conservation approaches; emphasizes proactive monitoring | Theoretical focus with minimal real-world examples; insufficient analysis of AI limitations |
| Ayoola et al. (2024) [106] | Biodiversity conservation | Big Data and AI | Highlights predictive modeling and habitat monitoring; integrates big data approaches | Overemphasis on the USA, limiting generalizability; lacks discussion on ethical and regulatory aspects |
| Jha et al. (2021) [110] | Urban planning | AI for smart cities | Broad review of AI's role in urban planning; discusses smart infrastructure and sustainability | Largely conceptual; lacks real-world case studies and algorithmic details |
| Sanchez et al. (2023) [111] | Urban planning | AI for sustainability | Focuses on AI's role in minimizing ecological footprints; forward-looking perspective | Limited focus on practical implementation; insufficient exploration of ethical/social concerns |
| Chen et al. (2021) [112] | $CO_2$ Emissions forecasting | AI for energy efficiency | Strong methodological framework; practical insights for emissions reduction | Limited scalability discussion; lacks analysis of computational constraints |
| Aras and Hanifi Van (2022) [113] | $CO_2$ emissions forecasting | AI for energy and emissions | Proposes an interpretable forecasting framework; holistic approach to energy-emissions link | Focuses on forecasting but lacks policy integration |
| Nazir et al. (2024) [114] | $CO_2$ Emissions forecasting | AI ensemble strategies | Uses advanced AI techniques for predictive analytics; informs policy decisions | Preprint status—findings not yet peer-reviewed; lacks discussion on ethical and computational aspects |

AI has played a crucial role in enhancing biodiversity conservation efforts, particularly in wildlife monitoring, ecosystem management, and habitat preservation. Raihan (2023) [104] provides an extensive overview of machine learning applications in conservation, demonstrating their role in optimizing resource management strategies. While the study highlights practical applications, it lacks empirical case studies that validate AI's effectiveness in real-world conservation settings. Addressing challenges related to AI adoption, data availability, and ethical concerns would strengthen its contribution to biodiversity science.

Ullah, Saqib, and Xiong (2024) [105] bridge traditional and AI-driven conservation approaches, emphasizing ecosystem monitoring and threat identification. The study effectively positions AI as a tool to complement classical biodiversity methods, but remains highly theoretical, lacking real-world applications and pilot projects. Further research should focus on practical implementation, incorporating interdisciplinary collaborations to address AI's algorithmic biases and interpretability concerns.

Ayoola et al. (2024) [106] examine AI's role in big data-driven biodiversity monitoring, particularly in habitat preservation and species tracking. This study underscores the benefits of predictive modeling for conservation planning, yet is heavily focused on the USA, limiting the generalizability of its findings to other ecological and socio-economic contexts. A more diverse dataset and regulatory discussion would enhance the study's applicability to global conservation efforts.

The use of AI in urban planning and smart city development has the potential to optimize infrastructure, reduce ecological footprints, and improve urban sustainability. Jha et al. (2021) [110] offer a broad conceptual review of AI's applications in urban design, particularly in smart infrastructure and indoor environment optimization. However, this study remains highly theoretical, lacking empirical validation and real-world case studies. Future research should focus on specific AI algorithms, such as biogeography-based optimization (BBO), which has demonstrated effectiveness in smart city planning.

Sanchez et al. (2023) [111] highlight AI's role in sustainable urban development, presenting a forward-looking perspective on how AI can minimize urban ecological footprints. While the study effectively discusses potential AI transformations, it lacks practical implementation strategies and case studies. Additionally, it overlooks critical ethical concerns, such as data privacy, urban bias, and accessibility in AI-driven planning decisions. Addressing these gaps through empirical studies and interdisciplinary frameworks would provide a more comprehensive understanding of AI's role in urban sustainability.

AI-driven approaches to $CO_2$ emissions modeling have shown promising results in enhancing energy efficiency and environmental policy formulation. Chen et al. (2021) [112] present a methodologically sound framework for AI-assisted energy and carbon footprint modeling, providing practical insights into emissions reduction. However, the study does not address scalability concerns or the computational costs of deploying AI models in large metropolitan regions. Including discussions on low-power AI models and edge computing applications could improve the study's applicability.

Aras and Hanifi Van (2022) [113] propose an interpretable forecasting framework for energy consumption and emissions modeling, offering a holistic view of energy–environment interactions. While the study's emphasis on interpretable AI is commendable, it focuses primarily on forecasting, lacking integration with urban policy planning and emissions reduction strategies. Future research should explore how AI-driven insights can directly inform and shape environmental policies.

Nazir et al. (2024) [114] explored AI ensemble strategies for $CO_2$ emissions forecasting, employing advanced deep learning methodologies. The study's strong predictive analytics approach makes it highly valuable for policymakers, yet its preprint status means that

its findings have not been fully validated. Additionally, it does not sufficiently explore the computational and ethical challenges of AI-driven emissions modeling, limiting its practical relevance. Addressing these technical and ethical considerations will be crucial for ensuring the responsible deployment of AI in emissions forecasting.

Despite AI's growing role in environmental monitoring and decision making, several challenges remain across all case studies. A common limitation is the lack of real-world empirical validation, with many studies relying primarily on theoretical discussions rather than pilot implementations. Additionally, ethical and regulatory considerations are often underexplored, particularly concerning AI bias, transparency, and governance frameworks.

Future research should emphasize interdisciplinary collaborations, integrating environmental science, AI development, and policy making to create more comprehensive and impactful AI-driven environmental solutions. Expanding pilot projects, open source AI frameworks, and adaptive learning models will be essential for ensuring scalability, fairness, and long-term sustainability in AI applications for environmental monitoring.

### *3.5. AI Techniques: Rationale and Suitability*

The selection of AI techniques in environmental monitoring depends on data characteristics, interpretability requirements, computational constraints, and the specific challenges of each application. The choice between supervised, unsupervised, and deep learning methods is guided by the nature of the dataset (labeled vs. unlabeled), the need for real-time processing, and the feasibility of deploying models on edge or cloud-based systems.

#### 3.5.1. Supervised Learning for Predictive Modeling

Supervised learning algorithms, such as random forests, support vector machines (SVMs), and gradient boosting are widely used in environmental monitoring due to their ability to learn complex relationships between labeled input data (e.g., sensor readings) and output categories (e.g., pollution levels). In water quality prediction, these models excel at mapping chemical composition data (e.g., pH, turbidity, dissolved oxygen) to contamination risk levels. The ability to train on historical pollution events allows for accurate forecasting, enabling proactive environmental management. Additionally, ensemble methods like Random Forests improve robustness by reducing overfitting, making them well suited for noisy environmental data.

#### 3.5.2. Unsupervised Learning for Anomaly Detection

When labeled data are scarce, unsupervised learning techniques become essential. K-means clustering, DBSCAN, and autoencoders are particularly useful for detecting anomalies in climate data, where sudden deviations in temperature, humidity, or atmospheric pressure may indicate early signs of extreme weather events. These models group similar data points, allowing for the identification of patterns that deviate from the norm. For example, K-means clustering can detect unexpected temperature spikes that may signal an impending heatwave, while autoencoders trained on normal air quality patterns can flag unusual pollutant concentrations as potential hazards.

#### 3.5.3. Deep Learning for Complex Feature Extraction

For applications requiring high-dimensional data processing, such as remote sensing and satellite imagery, deep learning techniques outperform traditional models. Convolutional neural networks (CNNs) are particularly effective in aerial image analysis, where they leverage hierarchical feature extraction to identify deforestation patterns, land-use changes, and water contamination from spectral images. CNN-based models trained on multispectral satellite imagery can differentiate between healthy and stressed vegetation, providing insights into drought impact and illegal deforestation activities. Similarly, re-

current neural networks (RNNs) and long short-term memory (LSTM) networks are well suited for time-series predictions, such as climate trend forecasting based on historical temperature and precipitation data.

#### 3.5.4. Balancing Interpretability and Computational Constraints

Model selection is not solely dictated by accuracy but also by interpretability and computational feasibility. Decision makers in environmental science often require transparent, explainable models to justify interventions. Decision trees and linear regression models provide straightforward rule-based outputs, making them preferable for regulatory compliance and policy making. Conversely, deep learning models, while highly accurate, are often seen as "black boxes", limiting their adoption in cases where interpretability is crucial.

Furthermore, computational constraints play a pivotal role in deployment feasibility. Edge computing applications, such as IoT-based real-time monitoring stations, require lightweight models due to power and processing limitations. Decision trees and logistic regression are commonly deployed on edge devices for rapid, low-power inference, whereas cloud-based architectures can accommodate more complex deep learning models, benefiting from scalable GPU resources.

#### 3.5.5. Hybrid Approaches for Enhanced Performance

Increasingly, hybrid models combining multiple AI techniques offer superior performance in environmental monitoring. For instance, an unsupervised clustering method can first detect anomalies in water quality data, which are then fed into a supervised classification model to predict contamination sources. Similarly, CNNs can extract spatial features from satellite images, while traditional machine learning models use these features for classification, ensuring a balance between interpretability and computational efficiency.

## 4. Leveraging IoT and AI for Improved Environmental Decision Making

The fusion of the Internet of Things (IoT) and artificial intelligence (AI) [130–132] has brought significant advancements to environmental sciences [133], enabling more efficient and data-driven decision-making processes. By leveraging the complementary capabilities of these technologies, integrated systems facilitate improved data collection, real-time analysis, and the generation of actionable insights. This convergence enhances the ability to tackle complex environmental challenges with greater accuracy and speed. This chapter examines the synergy between IoT and AI, details the workflow of integrated systems, and explores the advantages of real-time monitoring and responsive management strategies (Figure 2).

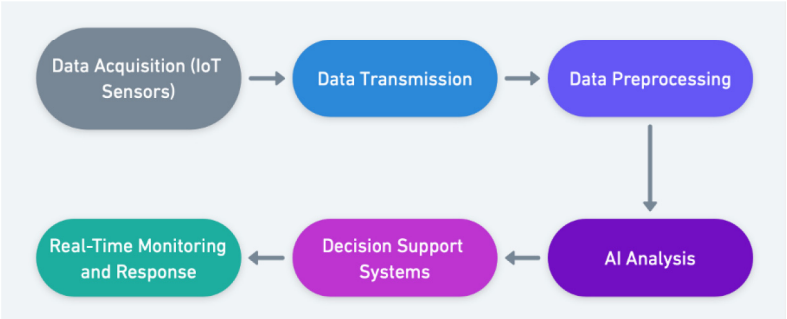



**Figure 2.** Integrated AI–IoT system workflow.

The combination of IoT and AI establishes a powerful framework for environmental monitoring and management [46,60,134]. IoT functions as the foundational layer, enabling the continuous and automated acquisition of extensive environmental data through a network of interconnected sensors and devices [135–137]. These sensors capture a diverse range of environmental parameters, including temperature [46,138], humidity [46], air and water quality [44,83], soil moisture [138–140], and pollutant concentrations, producing real-time data streams that offer a holistic perspective on environmental conditions (Table 3).

**Table 3.** Components of integrated AI–IoT Systems.

| Component | Description | Role in Integrated Systems |
| --- | --- | --- |
| IoT sensors [46,83,139] | Devices that collect environmental data such as temperature, humidity, pH, turbidity, etc. | Data acquisition from various environmental parameters |
| Data transmission [134,141,142] | Wireless communication standards (e.g., Wi-Fi, LoRaWAN, cellular networks) | Transmission of collected data to centralized repositories |
| Edge computing [143] | Local data processing units near data sources | Preliminary data analysis to reduce latency and bandwidth |
| AI algorithms [144–146] | Machine learning and deep learning models | Data analysis, pattern recognition, predictive modeling |
| Decision support systems (DSSs) [145,147] | Platforms that present data insights and recommendations to stakeholders | Facilitation of informed decision making |
| Visualization tools [148,149] | Dashboards and graphical interfaces | Presentation of complex data in an accessible format |

AI amplifies these capabilities by converting raw data from IoT devices into actionable insights. Through machine learning algorithms and advanced data analytics, it identifies patterns, detects anomalies, and predicts future environmental trends. This analytical capability allows AI models to execute advanced tasks such as predictive modeling, which forecasts pollution incidents or climate irregularities, and automated decision making, which selects the most effective responses using real-time data inputs. Consequently, the synergy between IoT and AI not only enhances data richness but also imbues the system with intelligent processing capabilities, making it possible to respond swiftly and accurately to emerging environmental challenges.

### *4.1. Data Quality Challenges and Preprocessing*

The accuracy of AI-driven environmental monitoring depends on high-quality input data. However, IoT sensors, satellite imagery, and monitoring stations often generate incomplete, noisy, and inconsistent data. Missing values arise from sensor failures and transmission errors, while sensor drift and human error introduce noise. Integrating diverse data sources—satellite imagery, IoT networks, and meteorological stations—requires harmonization to ensure consistency. Temporal and spatial gaps further disrupt data streams, complicating predictive modeling.

To address these issues, preprocessing techniques improve data reliability. Missing data imputation methods include mean, median, or mode replacement for minor gaps, time-series interpolation for continuous datasets, and machine learning-based imputation using k-nearest neighbors or autoencoders. Noise reduction techniques such as moving average smoothing and Kalman filters refine real-time sensor readings, while outlier detection

methods like isolation forest and DBSCAN identify anomalies in pollution and water quality data.

Normalization and standardization ensure consistency across datasets. Min-max scaling normalizes sensor readings such as temperature and pollution levels, while z-score normalization standardizes distributions. Feature engineering techniques like principal component analysis (PCA) reduce dimensionality in satellite images, while Fourier transform and wavelet analysis extract patterns from climate data.

Multi-source data integration is essential for accurate AI predictions. Data fusion techniques include low-level fusion, which merges raw sensor data, feature-level fusion, which combines extracted attributes from multiple datasets, and decision-level fusion, which aggregates independent AI model outputs. Spatial and temporal alignment ensures consistency, with geospatial harmonization linking satellite and ground-based sensor data, while time-series synchronization corrects inconsistencies in climate datasets. Ontology-based data standardization unifies formats across environmental domains, and edge computing allows local preprocessing before cloud-based AI analysis.

Effective data preprocessing enhances AI-driven environmental monitoring by improving accuracy, reducing noise, and ensuring interoperability. Future research should focus on automated preprocessing pipelines, self-learning imputation models, and AI-enhanced real-time data fusion for improved environmental prediction and decision making.

### *4.2. Workflow of Integrated Systems*

The integration of IoT and AI in environmental sciences follows a structured workflow, ensuring a smooth transition from data collection to decision making. The process begins with data acquisition, where IoT devices are strategically positioned to continuously gather environmental data. These sensors transmit information using communication protocols such as Wi-Fi, LoRaWAN, or cellular networks to centralized repositories or cloud platforms [136,150]. In some cases, edge computing is utilized to perform initial data processing near the source, minimizing latency and bandwidth consumption.

Once acquired, the data undergo preprocessing [151] and analysis to enhance its quality and usability for AI models. This stage includes data cleaning to eliminate noise and correct errors, ensuring accuracy and consistency. Data integration merges information from various sensors and sources to create a unified dataset, offering a comprehensive perspective on environmental conditions. Feature engineering extracts and selects the most relevant attributes that characterize the environmental phenomena under investigation [152].

Following preprocessing, AI model training applies machine learning and deep learning techniques to learn patterns from the refined dataset, enabling accurate predictions and decision making. Finally, real-time analytics processes incoming data streams instantaneously, generating immediate insights and facilitating proactive responses [153]. The final phase of the workflow involves decision support systems (DSSs) that leverage the outputs of AI models to aid stakeholders in making informed decisions [145]. Visualization tools present data and AI-driven insights through interactive dashboards, making complex information accessible and comprehensible [148]. Automated alerts and notifications are generated based on predefined thresholds or anomaly detection [60,154] patt, enabling swift responses to emerging issues. Additionally, DSSs provide actionable recommendations and strategic options derived from AI analyses, guiding policymakers, environmental managers, and other stakeholders in their decision-making processes. Feedback loops are established to incorporate the outcomes of decisions back into the system, continuously enhancing the accuracy and reliability of AI models through ongoing learning and adaptation.

### 4.3. Instantaneous Monitoring and Adaptive Response

One of the key benefits of combining IoT and AI in environmental sciences is the ability to enable real-time monitoring and rapid response [155–157]. This capability facilitates the immediate detection of environmental changes and the swift deployment of corrective measures, helping to minimize potential negative impacts. Real-time data processing offers multiple benefits that enhance the accuracy, responsiveness, and overall effectiveness of environmental oversight and control. It enables the immediate detection of environmental anomalies [158] or pollution events, allowing for timely interventions that can reduce the severity of environmental damage [159]. AI models can dynamically adapt their analyses and predictions based on the latest data, ensuring that insights remain accurate and relevant even as environmental conditions change. This adaptability supports enhanced responsiveness, enabling stakeholders to make agile decisions in the face of rapidly evolving environmental challenges [160–162]. Moreover, real-time processing optimizes operational efficiency by minimizing delays between data collection and action, thereby streamlining resource allocation and workflow management. Several rapid response scenarios illustrate the practical benefits of instantaneous monitoring and data-driven decision making. In flood management, IoT sensors placed in river basins and flood-prone areas collect data on water levels, rainfall, and soil saturation [163]. AI models process and interpret these data instantaneously to predict flood risks, triggering automated alerts and initiating evacuation plans or water diversion strategies to mitigate flood impacts [164,165]. Similarly, in urban air quality control, IoT sensors continuously monitor pollutant levels. When AI algorithms detect sudden spikes in pollutants like $PM_{2.5}$ or $NO_2$, immediate measures such as traffic restrictions or industrial emission controls can be implemented to safeguard public health [166,167]. Another example is wildfire detection [168,169], where IoT-enabled smoke detectors and temperature sensors deployed in forested areas provide real-time data on fire conditions [170]. AI agents analyze these data to identify early signs of wildfires, activating emergency response protocols and deploying firefighting resources swiftly to contain the blaze [169,171]. In water quality assurance, IoT sensors monitor parameters such as pH [172], dissolved oxygen [172,173], and pollutant concentrations in aquatic ecosystems [174–176]. AI models detect anomalies indicative of pollution events, prompting immediate remediation actions like halting industrial discharges, notifying relevant authorities, and initiating cleanup operations to protect water quality and aquatic life. Energy grid management also benefits from instantaneous tracking and AI integration [177]. IoT sensors track energy consumption and generation in real-time within smart grids [146,178,179]. AI systems process these data to enhance energy distribution, predict demand surges, and manage renewable energy sources efficiently, ensuring grid stability and reducing energy wastage [105]. The integration of IoT and AI fundamentally enhances decision-making processes in environmental sciences [180,181] by combining comprehensive data collection with intelligent analysis and actionable insights. The synergy between these technologies facilitates the development of resilient, instantaneous monitoring systems capable of adapting to dynamic environmental conditions and support proactive management strategies [60]. By streamlining the workflow from data acquisition to decision support, integrated AI–IoT systems empower environmental scientists, policymakers, and key stakeholders in tackling complex issues with greater efficacy and responsiveness. As technology evolves, the synergy between IoT and AI is set to propel further innovations, promoting more sustainable and resilient approaches to environmental management.

## 5. Environmental Data Focus

Environmental data function as the cornerstone for efficient oversight and informed decision making in environmental sciences [182,183]. Among the various types of environmental data, water quality and climate data are particularly crucial as a result of their profound impacts on ecosystems, public health [184,185], and global climate patterns [186,187]. This section examines the critical parameters tracked in these domains, the role of artificial intelligence (AI) in data analysis, and the Internet of Things (IoT) solutions utilized to enable comprehensive and continuous monitoring.

### *5.1. Water Quality Data*

Water quality monitoring [188,189] is crucial for preserving aquatic ecosystem health and safeguarding water resources for human consumption and industrial applications. Assessing water quality requires the evaluation of multiple parameters, which together offer a comprehensive understanding of the water's overall condition [175,190].

#### 5.1.1. Key Parameters Monitored

Monitoring water quality requires the continuous assessment of several critical parameters. The pH level reflects the acidity or alkalinity of the water [117,191], with variations from the neutral pH of 7 signaling potential contamination from sources like industrial effluents or acid rain, which may negatively impact aquatic ecosystems and water suitability [192]. Turbidity assesses the clarity of water by measuring the presence of suspended particles. High turbidity levels may result from soil erosion, algal blooms, or industrial effluents, which can harm aquatic organisms and complicate water treatment processes [193,194]. Dissolved oxygen (DO) is a crucial parameter that indicates the oxygen concentration in water [195,196], which is vital for the survival of fish and other aquatic organisms. Low DO levels can lead to hypoxic conditions, killing fish and disrupting aquatic ecosystems. Additionally, monitoring contaminants, including heavy metals [197,198] like lead and mercury, nutrients such as nitrates and phosphates, pathogens including bacteria and viruses [199,200], and various organic compounds, is crucial for ensuring water safety for consumption, recreation, and wildlife (Table 4).

**Table 4.** Key parameters monitored in water quality.

| Parameter | Significance | Common Measurement Methods |
|---|---|---|
| pH [192,201,202] | Indicates acidity or alkalinity; deviations signal pollution sources | pH meters, colorimetric tests |
| Turbidity [192,193,203,204] | Measures water clarity; high levels indicate suspended particles | Turbidity meters, nephelometers |
| Dissolved Oxygen [195,196] | Essential for aquatic life; low levels can lead to hypoxic conditions | DO meters, titration methods |
| Contaminants [197,200] | Includes heavy metals, nutrients, pathogens, organic compounds | Spectroscopy, chromatography, biosensors |

#### 5.1.2. AI Applications in Water Quality

Artificial intelligence has revolutionized water quality monitoring [205] by enhancing predictive capabilities and automating detection processes. Predictive modeling utilizes AI-powered algorithms to examine historical and real-time data, enabling the anticipation of potential pollution events [206]. By recognizing patterns and relationships within the data,

these models can predict contamination events like industrial spills or agricultural runoff, enabling proactive measures to reduce their impact [207,208]. Furthermore, AI facilitates automated detection and alerts by processing continuous streams of data from connected IoT devices [209,210]. Machine learning models have the capability to identify anomalies in parameters like pH or DO, automatically triggering notifications to environmental authorities and stakeholders. This enables swift response actions to contain and remediate pollution sources, thereby minimizing environmental and public health risks [211,212].

#### 5.1.3. IoT Devices for Water Monitoring

The deployment of IoT devices is integral to the comprehensive and continuous monitoring of water quality. Various types of sensors are strategically deployed to ensure robust data collection and real-time monitoring. pH sensors [213] are installed at multiple depths and locations within water bodies to continuously monitor acidity levels. Turbidity sensors are placed upstream and downstream of potential pollution sources to detect changes in water clarity [214]. Dissolved oxygen sensors are deployed in critical zones such as breeding grounds and habitats to ensure adequate oxygen levels for aquatic life [215,216]. Additionally, specialized contaminant sensors are positioned near industrial discharge points and agricultural runoff areas to detect specific pollutants like heavy metals or organic compounds [217,218]. Case studies highlight the impact of IoT-integrated water quality tracking systems. For instance, in the Great Lakes region, an integrated system employs IoT sensors across various points to collect data on pH, turbidity, dissolved oxygen, and contaminants. AI agents analyze these data in real-time, detecting pollution events and predicting future trends, which informs remediation strategies [103]. This system significantly improved the region's ability to manage water quality proactively, reducing the incidence of pollution-related issues. Similarly, smart irrigation systems in agricultural settings utilize IoT-enabled soil and water sensors to monitor irrigation water quality, ensuring that crops receive optimal water conditions. AI algorithms analyze sensor data to dynamically adjust irrigation practices, preventing nutrient runoff and minimizing water waste. This method improves crop yields while conserving water resources and minimizing environmental impact [103].

#### 5.1.4. Successful AI–IoT Integrations in Water Monitoring

AI–IoT systems have been successfully implemented across various environmental monitoring applications, demonstrating significant improvements in accuracy, efficiency, and real-time decision making. These integrations have been particularly effective in water quality assessment, pollution detection, and climate impact analysis. The following case studies highlight notable implementations of AI–IoT solutions in diverse environmental settings.

One notable application of AI–IoT integration is in real-time environmental monitoring, where sensor networks continuously collect data on water parameters such as temperature, pH, turbidity, and dissolved oxygen levels. AI-driven models process these data to detect anomalies and predict hazardous conditions. For instance, a study by Khan et al. (2024) [134] developed an IoT-based water quality monitoring system in smart cities, leveraging low-cost sensors and machine learning algorithms to forecast pollution events, enhancing urban water safety. Similarly, Ooko et al. (2024) [219] demonstrated the effectiveness of AI–IoT solutions in monitoring and predicting rural household air pollution in Africa, with potential applications for real-time water contamination tracking.

Beyond urban applications, AI–IoT has transformed smart agriculture by enabling the real-time monitoring of soil moisture, temperature, and nutrient levels. A study by Konar (2024) [220] introduced an AI–IoT framework for automated crop management

in urban farming, integrating predictive analytics to optimize irrigation and fertilization schedules [221]. These systems have improved water efficiency, reducing resource consumption while maximizing agricultural yields.

In remote and resource-limited regions, dynamic AI–IoT architectures have enhanced adaptability by enabling updatable AI models on ultra-low-power IoT devices. Alselek et al. (2024) [222] proposed a framework where AI models are dynamically updated within 5G-enabled IoT networks, eliminating the need for firmware modifications and enhancing long-term deployment feasibility. This innovation is particularly beneficial for water monitoring networks in disaster-prone areas, where AI models must adjust dynamically to evolving environmental conditions.

AI–IoT has also been instrumental in biodiversity assessment and conservation. Rathoure et al. (2024) [223] demonstrated how IoT sensor networks and deep learning models were used to monitor species diversity and habitat health in marine and freshwater ecosystems. These systems provided real-time ecological insights, enabling more effective conservation strategies [224].

### *5.2. Climate Data*

Climate data encompass a wide array of indicators that reflect the state and dynamics of the global climate system. The precise monitoring and analysis of climate data are crucial for comprehending climate change, forecasting future trends, and developing effective mitigation and adaptation strategies [225–227].

#### 5.2.1. Essential Climate Indicators

Several key climate indicators provide insights into various aspects of the climate system [228,229]. Temperature measurements, both surface and atmospheric, are critical for assessing global warming trends, heatwave occurrences, and their impacts on ecosystems and human populations. Humidity levels measure the amount of moisture in the air, influencing weather patterns, precipitation, and the formation of phenomena such as fog and thunderstorms. Precipitation data, including rainfall and snowfall, are vital for water resource management, agriculture, and flood forecasting. Wind patterns, which analyze wind speed and direction, are essential for understanding weather systems, ocean currents, and the dispersal of pollutants and aerosols (Table 5).

**Table 5.** Key climate indicators.

| Climate Indicator | Importance | Monitoring Methods |
| --- | --- | --- |
| Temperature [230,231] | Assesses global warming trends and heatwave impacts | Thermometers, satellite sensors, weather stations |
| Humidity [232,233] | Influences weather patterns and precipitation | Hygrometers, weather balloons, ground-based sensors |
| Precipitation [234,235] | Vital for water resource management and flood forecasting | Rain gauges, radar systems, satellite observations |
| Wind patterns [236] | Essential for understanding weather systems and pollutant dispersion | Anemometers, Doppler radar, weather satellites |

#### 5.2.2. AI in Climate Data Analysis

The role of AI in the analysis of climate data, enhancing the accuracy of climate models and identifying significant trends and anomalies is crucial [100]. Climate modeling and forecasting benefit from AI algorithms, particularly deep learning models, which simulate complex climate interactions and predict future climate scenarios with greater precision. These models incorporate vast datasets from historical records and real-time observations, improving the accuracy of predictions related to temperature changes, precipitation and extreme weather phenomena [237]. Moreover, AI techniques facilitate the identification of trends and anomalies within climate data. Machine learning algorithms process extensive climate datasets to identify long-term patterns and sudden irregularities, such as rapid ice melt in polar areas or unforeseen shifts in rainfall patterns [238]. These insights are vital for informing focused climate action initiatives and developing strategies to address emerging climate-related issues.

#### 5.2.3. IoT Solutions for Climate Monitoring

Role of IoT technologies in the comprehensive and continuous collection of climate data through a system of sensors and remote sensing devices is crucial [239]. A network of IoT-enabled weather stations is distributed across various regions to collect live data on temperature, humidity, precipitation, wind speed, and other climate indicators. These weather stations provide granular data that enhance the resolution and accuracy of climate models and forecasts [137]. Alongside weather stations based on ground [240], satellite imaging technologies [241,242] and drones equipped [243] with climate monitoring instruments gather data from remote and inaccessible regions. These devices measure atmospheric composition, sea surface temperatures, and ice coverage, helping to provide a more comprehensive understanding of global climate dynamics. The integration of IoT solutions with AI-driven analytics ensures that climate data are not only collected comprehensively but also efficiently analyzed, providing timely and actionable insights for climate scientists and policymakers [154]. Water quality and climate data are fundamental components of environmental monitoring, underpinning efforts to sustain ecosystems, protect public health, and address climate change. The integration of those technologies has significantly enhanced the capabilities to monitor these data streams comprehensively and in real time. AI-driven predictive modeling and automated detection systems improve the responsiveness and accuracy of environmental management practices, while IoT-enabled sensor networks ensure continuous and detailed data collection. Combined, AI and IoT enable more informed and prompt decision making, allowing for preventive measures to preserve water quality, reduce climate impacts, and support environmental sustainability.

## 6. Advantages of Combining IoT and AI in Environmental Sciences

The combination of the Internet of Things (IoT) and Artificial Intelligence (AI) in environmental sciences offers numerous advantages that greatly improve the effectiveness, efficiency, and sustainability of environmental monitoring and management practices [180]. This integration enhances the precision and accuracy of data collection and analysis, optimizes resource use, supports scalable and adaptable monitoring systems, and enables proactive environmental management [60,238]. These benefits are key to tackle complex environmental challenges and fostering sustainable practices (Table 6).

**Table 6.** Comparative analysis of benefits from AI–IoT integration.

| Benefit | Description | Impact on Environmental Sciences |
| --- | --- | --- |
| Enhanced accuracy and precision [211,244] | Improved data calibration, noise reduction, and high-resolution insights through AI-driven data processing | More reliable and detailed environmental assessments |
| Improved timeliness of data and decisions [245,246] | Real-time data processing and automated decision making facilitate swift responses to environmental changes | Faster and more effective interventions, reducing environmental impacts |
| Cost efficiency and resource optimization [247,248] | Automation reduces labor costs, AI optimizes energy and resource usage, and predictive maintenance minimizes operational expenses | More sustainable and economically viable environmental initiatives |
| Scalability and flexibility of monitoring systems [246] | IoT networks can be easily expanded and adapted to new parameters, supported by edge computing and modular design | Ability to handle increasing data volumes and diverse monitoring needs |
| Facilitating proactive environmental management [249] | Early warning systems, predictive maintenance, and strategic planning enabled by AI insights | Shift from reactive to proactive management, enhancing sustainability and resilience |

### *6.1. Enhanced Accuracy and Precision*

The combination of IoT and AI significantly enhances the accuracy and precision of environmental data gathering and analysis. IoT devices, outfitted with high-precision sensors, constantly collect detailed and comprehensive data on a range of environmental parameters [211]. These sensors track variables like temperature, humidity, air and water quality, soil moisture, and pollutant levels with exceptional precision [244]. AI algorithms analyze the massive volume of data produced by these sensors, uncovering complex patterns and correlations that might be missed by conventional analysis techniques. This sophisticated data processing ensures the accuracy and reliability of the measurements, greatly minimizing the margin of error in environmental evaluations. For instance, in water quality monitoring systems like those deployed in the Great Lakes region [103], IoT sensors collect live data on pH, turbidity, pollutants and dissolved oxygen levels [250]. AI agents analyze these data continuously, detecting even slight deviations that may indicate pollution events. This level of precision allows for early detection and prompt remediation actions and ensuring the safety of water resources.

### *6.2. Improved Timeliness of Data and Decisions*

Timeliness is a critical factor in environmental sciences, where delayed responses to emerging issues can exacerbate problems and lead to significant environmental and public health impacts. The integration of IoT technology and AI ensures that data are not only collected in real-time but also processed and analyzed swiftly, facilitating prompt decision making [246]. IoT devices transmit data continuously, and AI systems are designed to process this information instantaneously, eliminating delays between data acquisition and actionable insights [147]. The automated decision-making capabilities of AI further enhance timeliness by enabling systems to respond autonomously to specific environmental conditions. For instance, in managing urban air quality, AI algorithms process data from

IoT sensors that monitor pollutant levels in real-time [156]. When pollutant levels surpass safe thresholds, the AI system can autonomously initiate measures such as adjusting traffic signals, activating pollution control mechanisms, or issuing public health advisories [251]. This swift responsiveness ensures that interventions are timely and effective, mitigating adverse effects on public health and the environment.

### 6.3. Cost Efficiency and Resource Optimization

Integrating IoT and AI leads to significant cost savings and the optimal utilization of resources by streamlining processes, minimizing the need for manual intervention, and enhancing operational efficiencies [247]. Automated data collection and analysis minimize the reliance on extensive manual monitoring efforts, thereby lowering labor costs and reallocating human resources to more strategic tasks [248]. Additionally, AI-driven predictive maintenance of IoT infrastructure anticipates equipment failures and schedules timely maintenance, preventing costly downtimes and extending the lifespan of monitoring systems. Energy efficiency is another area where cost savings are realized. AI algorithms may optimize the operation of IoT technology, managing energy consumption effectively [252]. In smart irrigation systems, for instance, AI determines the precise amount of water needed based on the live data of soil moisture, reducing water and energy usage [81]. This targeted approach not only conserves valuable resources but also lowers operational costs for farmers, enhancing agricultural productivity and sustainability. Furthermore, AI systems facilitate the efficient allocation of resources by analyzing data to identify potential terrains of environmental risk or resource scarcity. This informed allocation ensures that limited resources such as water, energy, and funding are utilized where they are most needed, maximizing their impact and promoting sustainable environmental management practices [253].

### 6.4. Scalability and Flexibility of Monitoring Systems

A key advantage of combining IoT and AI is the unmatched scalability and flexibility it provides to environmental monitoring systems. IoT networks can be easily expanded by adding more sensors to cover larger geographical areas or to monitor additional environmental parameters [254]. This scalability ensures that monitoring systems can grow in tandem with expanding environmental monitoring needs without compromising performance. The modular design of IoT devices and AI algorithms allows for the customization and expansion based on specific environmental requirements [255]. For example, a monitoring system initially designed to track water quality can be seamlessly adapted to include additional sensors for detecting specific contaminants or to integrate climate data parameters as needed. This flexibility ensures that monitoring systems remain relevant and effective as environmental conditions and monitoring objectives evolve over time. Moreover, the marriage of edge computing with IoT devices enhances the scalability and flexibility of these systems [256]. By conducting initial data processing at the edge, close to the data source, the system lowers latency and reduces bandwidth consumption, allowing for quicker initial analyses and more efficient data handling [257]. This decentralized approach supports the deployment of large-scale sensor networks in diverse and remote locations, ensuring comprehensive and real-time environmental monitoring.

### 6.5. Facilitating Proactive Environmental Management

Perhaps the most transformative benefit of integrating IoT and AI is the facilitation of proactive environmental management [258]. This proactive approach shifts the focus from reactive responses to environmental issues to anticipating and preventing them before they escalate. Live data acquisition and forecasts enable stakeholders to identify future environmental hazards early and implement preventive measures to mitigate their influence [259].

Early warning systems exemplify proactive management by detecting the signs of potential environmental hazards such as floods, wildfires, or pollution events before they fully develop. For example, IoT sensors that monitor water levels and rainfall intensity in flood-prone areas provide real-time data that AI models analyze to predict flood risks [260]. This prediction allows automated alerts to be activated and evacuation plans or water diversion strategies are initiated, effectively reducing the severity of flood impacts [261]. In addition to emergency response, proactive environmental management encompasses strategic planning and dynamic resource management [258]. AI-driven analytics offer insights into long-term environmental trends, supporting the formulation of sustainable policies and strategies aimed at climate resilience and resource conservation. For example, in renewable energy management [262], AI systems optimize the distribution and storage of energy based on live consumption of data and predictive forecasts, ensuring efficient utilization of renewable resources and improving the stability of the network [263]. Furthermore, proactive management extends to infrastructure maintenance, where AI predicts maintenance needs based on data from IoT sensors, preventing infrastructure failures and ensuring the continuous operation of crucial environmental monitoring systems [157]. This foresight not only enhances the reliability of environmental management practices, but also promotes long-term sustainability and resilience to evolving environmental challenges [254]. The integration of these technologies into environmental sciences offers a suite of benefits that profoundly enhance environmental monitoring and management capabilities. Enhanced accuracy and precision ensure reliable and detailed data collection, while improved timeliness facilitates swift and informed decision making. Cost efficiency and resource optimization make environmental initiatives more sustainable and economically viable, and the scalability and adaptability of AI–IoT systems enable flexible and comprehensive monitoring across diverse environments. Most importantly, the shift towards proactive environmental management enables the anticipation and mitigation of environmental challenges, fostering a sustainable and resilient future. As IoT and AI technologies evolve, their combined integration will undoubtedly be crucial in tackling the complex and ever-changing environmental challenges of our era, fostering innovations that support sustainability, resilience, and data-driven and informed decision making.

## 7. Challenges and Considerations

There are many challenges and considerations despite the transformative adoption of IoT and AI technologies that must be addressed to ensure effective and responsible deployment. This section explores the main challenges related to data quality and reliability, system integration and interoperability, security and privacy issues, technical and infrastructure limitations, as well as ethical and regulatory concerns (Table 7) (Figure 3).

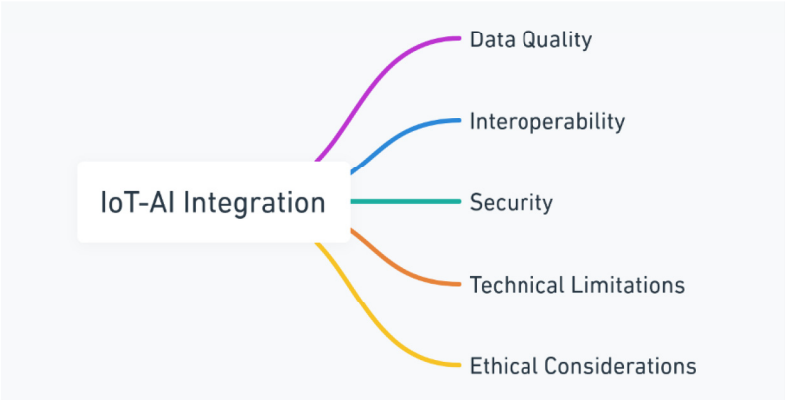



**Figure 3.** Overview of challenges in AI–IoT integration.

**Table 7.** Overview of challenges in AI–IoT integration.

| Challenge | Description | Potential Solutions |
| --- | --- | --- |
| Data quality and reliability [264–267] | The IoT sensors must provide accurate, consistent, and complete data; dealing with sensor drift and environmental interference | Regular calibration, robust data cleaning protocols, redundancy |
| Integration and interoperability [268–271] | Combining heterogeneous IoT devices and AI systems with diverse protocols and data formats; legacy system compatibility | Standardized protocols, middleware solutions, scalable frameworks |
| Security and privacy concerns [150,272–274] | Securing data from cyberattacks and unauthorized access; ensuring the protection of sensitive environmental and personal information. | Advanced encryption, authentication mechanisms, compliance with regulations |
| Technical and infrastructure limitations [221,275] | Reliable connectivity and power supply in remote areas; sufficient computational resources for AI models; maintenance of IoT infrastructure | Energy harvesting solutions, edge computing, robust maintenance plans |
| Ethical and regulatory considerations [224,276,277] | Tackling biases in AI models, ensuring transparency and interpretability, obtaining informed consent for data collection, and complying with data protection regulations | Fairness-aware algorithms, explainable AI techniques, clear governance frameworks |

### *7.1. Data Quality and Reliability*

The effectiveness of AI-powered environmental monitoring systems depends on the quality and dependability of the data gathered by IoT devices [266]. High-quality data ensure accurate analysis, reliable predictions, and trustworthy decision making, whereas poor data quality can lead to erroneous conclusions and ineffective interventions. Ensuring sensor accuracy and regular calibration is essential to maintain precise measurements. Over time, sensors may drift or degrade [264], necessitating ongoing maintenance to uphold data integrity. Additionally, data consistency and completeness are critical; incomplete or inconsistent data [267] can impair the performance of AI models, leading to inaccurate assessments. Environmental factors can also interfere with sensor performance [265], as IoT devices deployed in harsh or variable conditions may suffer from physical damage or exposure to extreme weather, thereby compromising data reliability. Robust data validation and cleaning protocols are necessary to identify and rectify anomalies, outliers, and errors in the dataset, maintaining the overall quality of the data being utilized.

### *7.2. Integration and Interoperability Issues*

Achieving the smooth integration of IoT and AI technologies across various platforms and systems is essential for developing unified environmental monitoring solutions. However, this integration is often hampered by the presence of heterogeneous devices, differing data formats, and varying communication protocols [268]. IoT devices typically utilize different communication standards, which can complicate the aggregation of data into a unified AI system [269]. Additionally, current environmental monitoring infrastructures may depend on legacy systems that are not naturally compatible with modern IoT and AI technologies [270], requiring careful planning and possible system upgrades

for seamless integration. The scalability of this integration also presents challenges; as the number of IoT devices increases, managing and consolidating data from a growing number of sensors becomes more complicated [271]. To enable effective AI model training and deployment, it is crucial to establish standardized data formats and ensure consistent data semantics across various IoT devices and platforms, facilitating comprehensive and coherent environmental analysis.

Interoperability Challenges

Interoperability between heterogeneous IoT devices and AI systems remains a key challenge in environmental monitoring, particularly in diverse environmental contexts where different sensor networks, communication protocols, and data formats must seamlessly integrate. Many environmental monitoring systems rely on IoT devices from multiple vendors, each using distinct data transmission standards, leading to compatibility issues that hinder AI-driven analysis.

To address these challenges, middleware platforms such as FIWARE [269] have been utilized to standardize data formats and APIs, enabling seamless communication across diverse IoT ecosystems. For example, in a multi-sensor flood monitoring network deployed in Southeast Asia, data from disparate sources—including pH sensors from Vendor A and turbidity sensors from Vendor B—were harmonized using JSON-LD schemas, ensuring that AI models could interpret and analyze the data consistently [165].

Infrastructure flexibility was further enhanced by deploying AI models within containerized environments, such as Docker, which facilitated seamless model execution across edge and cloud computing systems. This approach allowed real-time AI processing on IoT edge devices while ensuring scalability and cross-platform deployment when computational resources were available.

Despite these advancements, several interoperability challenges persist. Vendor-specific communication protocols, such as Zigbee and LoRaWAN, require adaptive gateways to enable cross-network integration, increasing system complexity. Furthermore, many environmental monitoring infrastructures, such as legacy wastewater treatment plants, operate on outdated SCADA (Supervisory Control and Data Acquisition) systems, which lack modern API support and require custom-built integration layers [275].

Future research should focus on developing unified IoT interoperability frameworks, integrating open source data exchange standards, and creating AI-driven protocol adaptation layers to enhance compatibility across legacy and next-generation IoT monitoring systems. By addressing these challenges, AI–IoT integration can become more scalable, adaptable, and effective in diverse environmental applications.

### *7.3. Security and Privacy Concerns*

The widespread use of IoT devices and the vast amount of data they collect present significant security and privacy challenges [272]. IoT devices are often placed in remote or unprotected locations, making them susceptible to both physical tampering and cyberattacks. It is crucial to implement strong encryption, authentication, and access control mechanisms to safeguard data integrity and prevent unauthorized access [273]. Furthermore, environmental data might inadvertently capture sensitive information about individuals or organizations, raising privacy concerns. For example, IoT sensors in agricultural fields could collect data that expose proprietary farming techniques, which could be exploited if not properly protected [274]. Additionally, AI-driven systems face various cybersecurity threats, such as data breaches, ransomware attacks [150], and adversarial machine learning attacks that can manipulate AI models to generate incorrect results. Compliance with data

protection regulations and the implementation of comprehensive security measures are vital to protect sensitive environmental data and preserve the trust of stakeholders.

### 7.4. Technical and Infrastructure Limitations

The deployment of integrated IoT and AI systems in environmental sciences is often constrained by technical and infrastructure limitations that can impede their effectiveness and scalability. Reliable Internet connectivity is crucial for the real-time transmission of data from IoT devices to AI systems [275]; however, in remote or underdeveloped areas, limited bandwidth and connectivity challenges can disrupt data flow and impair system performance [221]. Additionally, many IoT devices operate in environments where access to a stable power supply is challenging, relying instead on batteries or renewable energy sources, which necessitate efficient energy management to ensure continuous operation. AI models, especially deep learning-based ones, demand significant computational resources for both training and inference, which can be a limiting factor in environments with constrained resources. Deploying these models in such environments requires optimized algorithms and possibly edge computing solutions to handle the computational demands locally. Moreover, ensuring the ongoing maintenance and technical support of AI–IoT systems is critical for their sustained operation, involving the management of hardware failures, software updates, and addressing technical glitches promptly.

### 7.5. Ethical Considerations

The integration of IoT and AI in environmental sciences raises several ethical challenges that must be addressed to ensure responsible and equitable technology use. One major concern is algorithmic bias, where AI models inadvertently perpetuate biases from training data, potentially leading to unfair or discriminatory environmental management decisions [276]. Mitigating these biases requires implementing fairness-aware algorithms and ensuring the use of diverse and representative datasets.

Enhancing explainable AI (XAI) capabilities is crucial for ensuring accountability in decision-making processes, validating AI-driven recommendations, and building trust, as complex AI models, particularly deep learning systems, often function as "black boxes", making it difficult for stakeholders to understand how decisions are made.

Additionally, privacy concerns arise when AI–IoT systems collect environmental data that may indirectly involve human subjects. This includes data from agricultural landowners, industrial facilities, or community water sources. Informed consent and clear data usage policies must be established to protect individuals' rights and ensure ethical data collection practices [224].

Accountability is another critical factor. AI-driven decision-making in environmental monitoring, such as automated pollution detection and resource allocation, requires clear accountability mechanisms to address potential errors or unintended consequences. Establishing governance frameworks that define who is responsible for AI-driven actions will be crucial to ensure responsible deployment.

### 7.6. Regulatory Considerations

The evolving regulatory landscape governing AI and IoT in environmental sciences necessitates compliance with data privacy, security, environmental protection, and ethical AI usage laws [277]. Various international and regional regulations, such as the General Data Protection Regulation (GDPR) in Europe and industry-specific guidelines, impose stringent requirements on data collection, processing, and sharing.

Data protection laws emphasize the need for secure storage and handling of environmental data, particularly when IoT sensors monitor industrial emissions, water usage, or air quality, where data misuse could have legal and financial implications. Strong en-

cryption, anonymization, and secure access controls must be implemented to safeguard sensitive information.

Interoperability between different AI–IoT systems also requires standardized regulatory frameworks to ensure compatibility and security. Without clear regulations, inconsistencies in data formats, communication protocols, and compliance standards can hinder seamless integration across industries. Regulatory bodies are increasingly advocating for the establishment of common AI governance policies that ensure fairness, security, and accountability in automated decision-making systems.

Moreover, environmental regulations are incorporating AI-driven compliance monitoring tools to enforce pollution controls, sustainable resource usage, and emissions tracking. AI-powered compliance mechanisms must align with existing environmental laws and industry standards to avoid legal risks.

By proactively addressing these regulatory challenges through standardized protocols and collaborative policymaking, stakeholders can ensure the secure, transparent, and legally compliant deployment of AI–IoT systems [254]. Effective regulatory frameworks will not only mitigate risks but also foster innovation and trust in AI-driven environmental monitoring solutions.

### *7.7. Socioeconomic Factors in AI–IoT Adoption for Environmental Sciences*

The integration of AI and IoT in environmental monitoring is not solely a technological advancement; it is also influenced by socioeconomic factors that determine accessibility, scalability, and long-term sustainability. The successful deployment of these technologies depends on financial investment, economic feasibility, policy incentives, public perception, and workforce readiness.

The cost of deploying AI–IoT systems in environmental sciences remains a significant barrier, particularly for developing regions. High initial investment costs for sensor networks, cloud computing infrastructure, and AI model development can deter widespread adoption. While large corporations and government agencies may afford sophisticated AI-driven environmental monitoring systems, smaller municipalities and developing countries often lack the financial resources to implement such solutions at scale.

To address these challenges, cost-effective alternatives, such as edge computing and low-power IoT devices, are being explored to reduce dependency on expensive cloud-based processing. Additionally, public–private partnerships and government-funded initiatives can bridge financial gaps by subsidizing AI-driven environmental monitoring projects.

The acceptance of AI–IoT solutions in environmental sciences also depends on public trust and awareness. While AI-driven monitoring systems offer increased accuracy and efficiency, concerns about data privacy, job displacement, and decision-making transparency can create resistance to adoption.

For example, automated environmental monitoring systems may replace human-led inspections, raising concerns about job losses in traditional monitoring sectors. Additionally, communities may oppose AI-driven decision making if they perceive it as favoring corporate interests over public welfare. Addressing these concerns through transparent AI governance, participatory decision making, and public engagement campaigns can help build trust and ensure equitable access to AI-driven environmental solutions.

The widespread adoption of AI–IoT solutions in environmental sciences requires a skilled workforce capable of managing, maintaining, and interpreting AI-driven insights. However, many regions face a shortage of professionals with expertise in AI, machine learning, IoT integration, and environmental data analytics.

Investment in educational programs, vocational training, and interdisciplinary research initiatives can help bridge these skill gaps. Collaboration between universities,

industries, and government agencies can further equip environmental scientists with AI literacy while encouraging AI professionals to specialize in environmental applications.

While AI–IoT solutions hold promise for sustainable environmental management, there is a risk that their benefits will be unevenly distributed. High-income countries and technologically advanced industries may leverage AI for superior environmental monitoring, while low-income regions remain vulnerable to climate change and pollution due to lack of resources.

Ensuring equitable access to AI–IoT technologies requires open source initiatives, technology-sharing agreements, and funding models that enable developing nations to adopt low-cost AI-driven environmental monitoring systems. By making these technologies more inclusive, AI and IoT can contribute not only to environmental sustainability but also to global economic and social resilience.

## 8. Future Directions and Emerging Trends

The precision and efficiency of monitoring and decision-making processes have already seen significant improvements in environmental sciences thanks to the contributions of technologies like the Internet of Things (IoT) and artificial intelligence (AI). However, the technological landscape continues to evolve rapidly, promising even greater advancements and innovative applications in the future. This section explores the anticipated future directions and emerging trends that are poised to further revolutionize environmental sciences, with a particular emphasis on water quality and climate data (Table 8).

**Table 8.** Emerging trends in AI and IoT technologies for environmental sciences.

| Emerging Trend | Description | Potential Applications | Anticipated Impact |
|---|---|---|---|
| Advanced machine learning techniques [266,278–280] | Improvements in deep learning, transfer learning, federated learning, and reinforcement learning for more sophisticated environmental data analysis | Enhanced climate modeling, adaptive resource management | Greater predictive accuracy and adaptability |
| Evolution of IoT technologies [221,274–276,281] | Development of LPWAN, advanced sensors, energy harvesting, miniaturization, and edge computing for more robust and efficient environmental monitoring systems | Large-scale sensor networks, remote and real-time monitoring | Increased data granularity and operational efficiency |
| Integration with blockchain and edge Computing [224,238,254,277] | Utilizing blockchain for data integrity and traceability, and edge computing for decentralized processing | Secure data sharing, real-time analytics | Enhanced security, reduced latency, and improved scalability |
| Global environmental monitoring networks [278–280] | Establishment of standardized global networks for synchronized data collection and analysis | Comprehensive climate monitoring, disaster response coordination | Unified global efforts in addressing environmental challenges |
| Community engagement and citizen science [281–285] | Empowering communities through citizen-driven data collection, educational programs, interactive platforms, and collaborative research opportunities | Expanded data collection, enhanced public awareness | Democratized environmental monitoring and increased public participation |

### *8.1. Advancements in AI and Machine Learning Techniques*

Machine learning algorithms are constantly evolving, providing increasingly advanced tools and methodologies that can be applied to environmental sciences [63,286].

Advancements in deep learning architectures, including convolutional neural networks and recurrent neural networks, are enhancing the accuracy of pattern recognition and predictive modeling, enabling more precise outcomes in various fields. These advancements allow for the better interpretation of complex environmental data, including satellite imagery for climate monitoring and sensor data for water quality assessment [280,287]. Transfer learning and domain adaptation are emerging as crucial techniques, enabling AI models trained on particular data to be adapted for use in different but related domains. This is particularly beneficial in environmental sciences, where data scarcity in certain regions or parameters can be addressed by leveraging existing models trained on more abundant datasets [288]. Explainable AI (XAI) is another significant advancement, providing insights into how AI models make predictions [289]. This transparency of the AI system is crucial for ensuring trust and ensuring accountability in environmental decision making [290]. Gaining traction as a decentralized machine learning approach, federated learning [291] enables models to be trained across multiple devices or servers that hold local data samples, without the need to exchange the data itself [292]. This enhances data privacy and security, which is crucial for sensitive environmental data collected from various sources. Additionally, reinforcement learning is being explored for dynamic and adaptive environmental management tasks, such as optimizing water distribution [279] in smart irrigation systems or managing renewable energy [278] resources within smart grids.

### *8.2. Evolution of IoT Technologies for Environmental Applications*

Long-range communication with minimal energy consumption, provided by low-power wide-area networks (LPWAN) [293], which includes LoRaWAN [274,276] and NB-IoT [221,275] makes these technologies ideal for deploying large-scale sensor networks in remote or hard-to-reach areas, offering robust, efficient, and versatile solutions for environmental monitoring. Advanced sensor technologies are also being developed, enhancing the sensitivity, accuracy, and multifunctionality of environmental sensors. Innovations such as biosensors for detecting biological contaminants in water and advanced meteorological sensors for detailed climate data collection are expanding the scope of environmental monitoring [281]. Energy harvesting and sustainable power solutions are addressing one of the critical challenges of IoT deployment—power supply. Advances in solar, wind, and kinetic energy harvesting enable IoT devices to operate sustainably without relying on traditional power sources. This is particularly beneficial for long-term deployments in environmentally sensitive or remote locations [294]. Additionally, the trend toward the miniaturization and integration of IoT devices [255] allows for the deployment of dense sensor networks with minimal environmental footprint. This enables detailed and extensive environmental monitoring, which is crucial for detecting complex spatial and temporal variations in water quality and climate data. Edge computing is also becoming integral to IoT technologies, allowing for preliminary data processing and analysis at the source. This reduces latency and bandwidth usage, enabling quicker initial analyses and more efficient data handling. The integration of edge computing with IoT enhances the scalability and flexibility of environmental monitoring systems, ensuring that they can adapt to growing data volumes and diverse environmental conditions.

### *8.3. Integration with Other Technologies*

The combination of IoT and AI with emerging technologies is set to significantly boost their impact on environmental sciences. A notable example is blockchain technology, which can ensure the integrity and transparency of environmental data collected by IoT devices [254]. Through its decentralized ledger, blockchain facilitates the secure sharing of data among various stakeholders, promoting trust and accountability in environmental

monitoring and reporting. Furthermore, blockchain-based smart contracts can automate and enforce environmental policies and agreements, using data from IoT devices and AI analyses to ensure adherence and improve resource management efficiency.

As discussed earlier, edge computing enhances the functionality of AI–IoT systems by enabling decentralized data processing. This integration allows for real-time analytics and decision making to occur closer to the source of the data, reducing reliance on centralized cloud systems and increasing resilience. Additionally, augmented reality (AR) and virtual reality (VR) technologies are being combined with IoT and AI [224,295] to provide immersive visualizations of complex environmental data. These technologies make it easier for stakeholders to engage with and interpret environmental conditions and trends, thus supporting more effective decision making.

Although still in its early stages, quantum computing presents the potential to transform environmental modeling and simulation [296]. Quantum algorithms are capable of solving complex environmental problems much more rapidly than traditional computers, offering the possibility of more precise climate models, the better optimization of renewable energy systems, and large-scale environmental simulations [238,277]. As quantum computing continues to evolve, its integration with IoT and AI could lead to groundbreaking advancements in the field of environmental sciences.

### *8.4. Potential for Global Environmental Monitoring Networks*

The establishment of global environmental monitoring networks powered by IoT and AI promises comprehensive and synchronized data collection and analysis on a planetary scale. Developing standardized protocols and data formats is essential for integrating diverse sensor networks and ensuring seamless data exchange across different regions and countries [279]. International collaboration is crucial to establish common standards that facilitate global monitoring efforts, enabling the aggregation of data from various sources into centralized or federated data repositories [280]. Collaborative platforms that enable data sharing and collaboration among international environmental agencies, research institutions, and governments can enhance the effectiveness of global monitoring efforts [278]. These platforms enable the sharing of best practices, technologies, and insights derived from AI-driven analyses, fostering a collaborative, unified approach to tackling global environmental challenges. Through real-time global dashboards that visualize environmental data alongside AI-generated insights, stakeholders gain immediate access to vital information, which in turn supports worldwide efforts to address critical issues like climate change, water scarcity, and other urgent environmental concerns.

Additionally, global monitoring networks play an essential role in coordinating disaster response efforts. By analyzing data from various regions, AI algorithms can predict the spread and potential impact of natural disasters, helping to streamline and enhance emergency response actions. This comprehensive approach ensures the efficient allocation of resources and ensures that communities are better prepared to handle and mitigate the effects of environmental crises.

### *8.5. Enhancing Community Engagement and Citizen Science*

Empowering communities and involving citizens in environmental monitoring can significantly change the adoption power and reach of IoT and AI technologies. Citizen-driven data collection expands the scope and granularity of environmental data, providing valuable insights from diverse locations [285]. Initiatives that encourage individuals to deploy IoT devices and contribute data to environmental monitoring projects can leverage smartphones, wearable sensors, and DIY IoT kits. This could enable widespread participation in data collection, fostering a sense of ownership and accountability among community

members. Educational programs and workshops that provide training and resources on how to use IoT and AI tools for environmental monitoring are essential for fostering greater participation and engagement. These programs equip citizens with the skills needed to deploy sensors, interpret data, and contribute to AI-driven analyses, thereby enhancing the overall effectiveness of environmental monitoring systems [284]. Engaging platforms and mobile apps that enable citizens to view, explore, and interact with environmental data can significantly boost involvement [283]. These platforms provide real-time updates, personalized insights, and opportunities for community-driven initiatives, making environmental data more accessible and actionable for the general public. Feedback systems that enable citizens to share observations, propose improvements, and contribute to decision-making processes help ensure that environmental monitoring systems address community needs and perspectives [281]. Incentivizing participation through recognition programs, rewards, or contributions to community projects can motivate greater involvement in citizen science initiatives. This increased participation leads to more extensive data collection and fosters a culture of environmental stewardship. Collaborative research opportunities that involve citizens in data analysis and solution development bridge the gap between scientific communities and the public, promoting a deeper understanding of environmental issues and the co-creation of innovative solutions. The future of environmental sciences is set to be profoundly influenced by the ongoing advancements and emerging trends in AI and IoT technologies. As AI and machine learning techniques continue to progress, they will improve the analytical capabilities and forecasting abilities of environmental monitoring systems. At the same time, the development of IoT technologies will offer more robust, efficient, and adaptable data collection methods. The incorporation of additional technologies, such as blockchain and edge computing, will enhance both the functionality and security of these systems. Moreover, the potential creation of global environmental monitoring networks holds the promise of comprehensive and coordinated data collection on a worldwide scale.

Fostering greater community involvement and supporting citizen science initiatives will democratize environmental monitoring [282], broadening the impact and reach of IoT and AI technologies. As these trends advance, the integrated use of IoT and AI will be central to promoting environmental sustainability, resilience, and data-driven decision making. Embracing these emerging directions will empower environmental scientists, policymakers, and communities to work together in addressing the urgent environmental challenges of our time, leading to a more sustainable and resilient future.

### *8.6. Research Gaps and Future Directions*

Despite significant advancements in AI–IoT integration for environmental monitoring, several critical gaps remain, limiting the scalability, reliability, and societal impact of these technologies. Addressing these challenges will be essential for ensuring the long-term effectiveness of AI-driven environmental solutions.

One major gap is long-term system sustainability. While numerous studies demonstrate successful short-term AI–IoT deployments, longitudinal data on sensor drift, model degradation, and hardware resilience remain scarce. Most AI models are trained on static datasets, yet real-world environmental monitoring systems require adaptive mechanisms to detect and correct sensor inaccuracies over extended periods [264]. Future research should explore self-healing AI architectures, incorporating automated recalibration techniques to maintain accuracy in evolving environmental conditions.

Another key challenge is real-time explainability. While methods like SHAP (Shapley Additive Explanations) provide transparency for static AI predictions, explainable AI (XAI) techniques for streaming IoT data remain underdeveloped [290]. Environmental

monitoring applications require dynamic interpretability frameworks that can clarify why AI models issue real-time alerts or classify specific pollution events. Edge-based XAI approaches could improve on-device decision-making by providing interpretable insights at the point of data collection.

Socio-technical governance is another underexplored area, particularly in balancing data privacy regulations (e.g., GDPR) with public interest in pollution transparency [224]. While the AI–IoT systems generate large-scale environmental data, few governance frameworks address how regulatory compliance, community engagement, and ethical considerations should be integrated into AI deployment strategies. Future work should prioritize participatory design approaches, ensuring that environmental monitoring technologies are developed collaboratively with affected communities, policymakers, and industry stakeholders.

To address these gaps, future efforts should focus on three key priorities: (1) federated learning for privacy-preserving AI collaborations, enabling decentralized model training without exposing raw environmental data; (2) Digital Twins for scenario testing, allowing researchers to simulate AI–IoT system performance in diverse environmental conditions before large-scale deployment; and (3) inclusive AI frameworks that incorporate perspectives from marginalized communities, ensuring that environmental monitoring systems equitably benefit all populations. By addressing these challenges, AI–IoT solutions can become more resilient, transparent, and socially responsible, supporting sustainable environmental management on a global scale.

## 9. Comparative Analysis with Existing Reviews

Several recent studies have explored the integration of AI and IoT for environmental monitoring, focusing on water quality and climate data. While these reviews contribute valuable insights, the present study distinguishes itself by offering a broader perspective that includes methodological advancements, interdisciplinary applications, and a structured evaluation of case studies. The following comparative analysis highlights key differences and novel contributions.

*Novel Contributions of This Study*

Unlike Popescu (2024) [60], which primarily examines machine learning for pollution monitoring, this study provides a structured analysis of AI–IoT applications across multiple environmental domains. It expands on real-time AI–IoT integration, addressing preprocessing challenges, data harmonization, and multi-source sensor fusion, aspects that are only partially covered in existing reviews.

Khan (2024) [134] focuses on real-time AI-driven trend prediction using GRU-Autoencoder, demonstrating adaptive AI techniques for environmental monitoring. However, this study goes further by evaluating long-term ecosystem monitoring and discussing AI model scalability. Unlike Khan's work, which is limited to short-term data predictions, this study incorporates historical climate analysis and ecosystem-based AI modeling, offering a comprehensive approach to environmental management.

Arabelli (2024) [297] presents case studies on AI–IoT efficiency in environmental change detection, but does not explore ethical concerns, governance frameworks, or regulatory compliance. In contrast, this study examines data governance, privacy concerns, and explainable AI (XAI) for environmental monitoring, ensuring that AI-driven decision making aligns with regulatory standards.

Arowolo (2024) [298] emphasizes AI–IoT applications for long-term ecosystem analysis, but lacks an in-depth evaluation of predictive models and data scalability. This study bridges that gap by presenting a comparative analysis of AI models, exploring hybrid ap-

proaches that combine deep learning with traditional predictive analytics, making AI-based environmental monitoring more robust and adaptable.

Panduman (2024) [155] introduces the SEMAR platform, focusing on AI in IoT applications. While it presents a system-level approach, it does not discuss interoperability challenges, multi-source data harmonization, or environmental-specific AI applications. This study offers a broader perspective on AI techniques tailored to environmental challenges, ensuring seamless data integration from heterogeneous sources.

Manongga (2024) [299] proposes the AIKU model for real-time air quality monitoring, demonstrating high accuracy and scalability. However, it remains focused on air pollution and wireless sensor networks, without expanding into water quality, climate monitoring, or multi-environmental AI applications. This study extends beyond air monitoring by addressing AI applications in hydrological and meteorological contexts, providing a more comprehensive environmental perspective.

This comparative analysis highlights the novel contributions of this study, particularly in AI–IoT integration for water quality and climate monitoring, case study-based evaluations, preprocessing techniques, and ethical considerations. Unlike existing reviews, which focus on specific AI applications, this study presents a holistic approach, discussing AI model comparisons, multi-source data integration, and regulatory compliance. Future research could further explore explainable AI (XAI) for environmental monitoring, as well as low-resource AI deployment for regions with limited computational infrastructure.

## 10. Conclusions

The integration of artificial intelligence (AI) agents with the Internet of Things (IoT) marks a transformative advancement in environmental sciences, offering exceptional capabilities for monitoring, analyzing, and managing environmental data. This article has highlighted the multifaceted applications of AI and IoT in tackling critical environmental challenges, particularly focusing on water quality and climate data. These technologies synergistically enhance the precision and reliability of environmental monitoring while facilitating informed decision making, setting the stage for sustainable and resilient environmental management practices.

Comprehensive monitoring and data collection are enabled through IoT sensors deployed in diverse environmental contexts. These sensors continuously gather granular data, such as pH values, turbidity levels, dissolved oxygen concentrations in aquatic systems, as well as climate parameters like temperature, humidity, precipitation, and wind patterns. The substantial volume and diversity of data provided by these devices form a robust foundation for AI-driven analysis, ensuring real-time access to high-quality information for environmental scientists.

AI agents employ advanced machine learning and deep learning techniques to process and interpret the data collected by IoT devices. Through predictive modeling and automated detection systems, these technologies anticipate pollution events, climate anomalies, and other environmental shifts. This capability empowers stakeholders to take proactive measures, minimizing adverse impacts and enhancing the overall efficacy of environmental management strategies.

Operational efficiency and cost optimization are other benefits of AI and IoT integration. Automating data collection and analysis reduces dependency on manual efforts, minimizes operational costs, and ensures the efficient allocation of resources. Moreover, the AI-powered predictive maintenance of IoT infrastructure prevents costly equipment failures and prolongs the lifespan of monitoring systems.

The scalability and flexibility of AI–IoT systems make them adaptable to evolving environmental conditions and emerging challenges. These systems can expand sensor

networks and incorporate new parameters as required, maintaining their effectiveness and relevance over time. This adaptability allows for proactive environmental management, shifting the focus from reactive responses to strategic, forward-thinking solutions. Real-time data processing supports early warning systems, dynamic resource management, and informed planning, promoting sustainability and resilience.

Despite their significant advantages, the integration of AI and IoT in environmental sciences poses challenges such as ensuring data quality, achieving seamless integration, safeguarding security and privacy, addressing technical limitations, and managing ethical and regulatory concerns. Overcoming these challenges requires the development of robust frameworks, standardized protocols, and collaborative initiatives.

Looking forward, advancements in AI and IoT technologies, combined with innovations like blockchain and edge computing, hold immense potential to enhance environmental monitoring and management. Establishing global environmental monitoring networks and engaging communities through citizen science initiatives will democratize data collection, fostering inclusive approaches to environmental stewardship. As these technologies evolve, their synergistic application will play a pivotal role in addressing pressing environmental issues, driving innovations that support sustainability, resilience, and informed decision making.

The convergence of AI and IoT is a significant milestone in environmental sciences. By leveraging these technologies, environmentalists and data scientists can develop strategies to monitor and protect vital resources. As we navigate complex environmental challenges, the integrated AI–IoT framework emerges as a cornerstone for advancing sustainable and resilient management practices. Embracing these innovations equips us with the tools to anticipate and mitigate future challenges, ensuring a healthier, more sustainable world for future generations.

**Author Contributions:** Conceptualization, T.M. and I.D.; methodology, T.M. and I.D.; investigation, T.M., I.D., E.K., P.K., A.Ł., S.S. and A.N.; resources, T.M., I.D., E.K., P.K., A.Ł., S.S. and A.N.; writing—original draft preparation, T.M., I.D. and E.K.; writing—review and editing, T.M., I.D., E.K., P.K., A.Ł., S.S. and A.N.; visualization, T.M. 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.

**Conflicts of Interest:** Author Sylwia Sokołowska was employed by Polish Society of Bioinformatics and Data Science BioData. 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. Brauch, H.G. Introduction: Globalization and Environmental Challenges: Reconceptualizing Security in the 21st Century. In *Globalization and Environmental Challenges*; Brauch, H.G., Spring, Ú.O., Mesjasz, C., Grin, J., Dunay, P., Behera, N.C., Chourou, B., Kameri-Mbote, P., Liotta, P.H., Eds.; Hexagon Series on Human and Environmental Security and Peace; Springer: Berlin/Heidelberg, Germany, 2008; Volume 3, pp. 27–43, ISBN 978-3-540-75976-8.
2. Crompton, T.; Kasser, T. *Meeting Environmental Challenges: The Role of Human Identity*; WWF-UK: Surrey, UK, 2009; ISBN 978-1-900322-64-5.
3. Conca, K.; Dabelko, G.D. (Eds.) *Green Planet Blues: Critical Perspectives on Global Environmental Politics*, 6th ed.; Taylor & Francis Group: New York, NY, USA; London, UK, 2020; ISBN 978-0-429-32220-4.

4. Arthington, A.H.; Bunn, S.E.; Poff, N.L.; Naiman, R.J. The Challenge of Providing Environmental Flow Rules to Sustain River Ecosystems. *Ecol. Appl.* **2006**, *16*, 1311–1318. [CrossRef]
5. Palomino, J.; Muellerklein, O.C.; Kelly, M. A Review of the Emergent Ecosystem of Collaborative Geospatial Tools for Addressing Environmental Challenges. *Comput. Environ. Urban Syst.* **2017**, *65*, 79–92. [CrossRef]
6. Hu, Y.; Cheng, H.; Tao, S. Environmental and Human Health Challenges of Industrial Livestock and Poultry Farming in China and Their Mitigation. *Environ. Int.* **2017**, *107*, 111–130. [CrossRef] [PubMed]
7. Mwangi, W.; De Figueiredo, P.; Criscitiello, M.F. One Health: Addressing Global Challenges at the Nexus of Human, Animal, and Environmental Health. *PLoS Pathog.* **2016**, *12*, e1005731. [CrossRef] [PubMed]
8. Mellor, J.W. The Intertwining of Environmental Problems Poverty. *Environ. Sci. Policy Sustain. Dev.* **1988**, *30*, 6–30. [CrossRef]
9. Mabogunje, A.L. Poverty and Environmental Degradation: Challenges within the Global Economy. *Environ. Sci. Policy Sustain. Dev.* **2002**, *44*, 8–19. [CrossRef]
10. Masron, T.A.; Subramaniam, Y. Does Poverty Cause Environmental Degradation? Evidence from Developing Countries. *J. Poverty* **2019**, *23*, 44–64. [CrossRef]
11. Patz, J.A.; Frumkin, H.; Holloway, T.; Vimont, D.J.; Haines, A. Climate Change: Challenges and Opportunities for Global Health. *JAMA* **2014**, *312*, 1565. [CrossRef] [PubMed]
12. Solecki, W. Urban Environmental Challenges and Climate Change Action in New York City. *Environ. Urban.* **2012**, *24*, 557–573. [CrossRef]
13. Roberts, D.; O'Donoghue, S. Urban Environmental Challenges and Climate Change Action in Durban, South Africa. *Environ. Urban.* **2013**, *25*, 299–319. [CrossRef]
14. Zhou, Y.; Khu, S.-T.; Xi, B.; Su, J.; Hao, F.; Wu, J.; Huo, S. Status and Challenges of Water Pollution Problems in China: Learning from the European Experience. *Environ. Earth Sci.* **2014**, *72*, 1243–1254. [CrossRef]
15. Han, D.; Currell, M.J.; Cao, G. Deep Challenges for China's War on Water Pollution. *Environ. Pollut.* **2016**, *218*, 1222–1233. [CrossRef] [PubMed]
16. Tang, W.; Pei, Y.; Zheng, H.; Zhao, Y.; Shu, L.; Zhang, H. Twenty Years of China's Water Pollution Control: Experiences and Challenges. *Chemosphere* **2022**, *295*, 133875. [CrossRef] [PubMed]
17. Carr, D.L.; Suter, L.; Barbieri, A. Population Dynamics and Tropical Deforestation: State of the Debate and Conceptual Challenges. *Popul. Environ.* **2005**, *27*, 89–113. [CrossRef] [PubMed]
18. Haigh, M.J.; Jansky, L.; Hellin, J. Headwater Deforestation: A Challenge for Environmental Management. *Glob. Environ. Chang.* **2004**, *14*, 51–61. [CrossRef]
19. Kumar, R.; Kumar, A.; Saikia, P. Deforestation and Forests Degradation Impacts on the Environment. In *Environmental Degradation: Challenges and Strategies for Mitigation*; Singh, V.P., Yadav, S., Yadav, K.K., Yadava, R.N., Eds.; Water Science and Technology Library; Springer International Publishing: Cham, Switzerland, 2022; Volume 104, pp. 19–46, ISBN 978-3-030-95541-0.
20. Rands, M.R.W.; Adams, W.M.; Bennun, L.; Butchart, S.H.M.; Clements, A.; Coomes, D.; Entwistle, A.; Hodge, I.; Kapos, V.; Scharlemann, J.P.W.; et al. Biodiversity Conservation: Challenges Beyond 2010. *Science* **2010**, *329*, 1298–1303. [CrossRef]
21. Arora, N.K.; Fatima, T.; Mishra, I.; Verma, M.; Mishra, J.; Mishra, V. Environmental Sustainability: Challenges and Viable Solutions. *Environ. Sustain.* **2018**, *1*, 309–340. [CrossRef]
22. Kim, J.W. The Environmental Impact of Industrialization in East Asia and Strategies toward Sustainable Development. *Sustain. Sci.* **2006**, *1*, 107–114. [CrossRef]
23. Perreault, T.A.; Bridge, G.; McCarthy, J. (Eds.) *The Routledge Handbook of Political Ecology*; Routledge International Handbooks; Routledge: London, UK; Taylor & Francis Group: New York, NY, USA, 2015; ISBN 978-1-138-79433-7.
24. Patnaik, R. Impact of Industrialization on Environment and Sustainable Solutions–Reflections from a South Indian Region. *IOP Conf. Ser. Earth Environ. Sci.* **2018**, *120*, 012016. [CrossRef]
25. Keiner, M.; Koll-Schretzenmayr, M.; Schmid, W.A. *Managing Urban Futures: Sustainability and Urban Growth in Developing Countries*; Routledge: London, UK, 2016; ISBN 978-1-315-24982-7.
26. Dodman, D. Environment and Urbanization. In *International Encyclopedia of Geography*; Richardson, D., Castree, N., Goodchild, M.F., Kobayashi, A., Liu, W., Marston, R.A., Eds.; Wiley: Hoboken, NJ, USA, 2017; pp. 1–9, ISBN 978-0-470-65963-2.
27. Akash, M.; Akter, J.; Tamanna, T.; Kabir, M.R. The Urbanization and Environmental Challenges in Dhaka City. *SSRN Electron. J.* **2018**, 145–157. [CrossRef]
28. Iyiola, A.O.; Akinsorotan, O.A.; Ojeleye, A.E.; Fajimolu, A.O. An Overview of Environmental Resources in Africa: Emerging Issues and Sustainable Exploitation. In *Sustainable Utilization and Conservation of Africa's Biological Resources and Environment*; Izah, S.C., Ogwu, M.C., Eds.; Sustainable Development and Biodiversity; Springer: Singapore, 2023; Volume 32, pp. 543–570, ISBN 978-981-19-6973-7.
29. Numbere, A.O.; Maduike, E.M. The Impact of Unsustainable Exploitation of Forest and Aquatic Resources of the Niger Delta, Nigeria. In *Biodiversity in Africa: Potentials, Threats and Conservation*; Chibueze Izah, S., Ed.; Sustainable Development and Biodiversity; Springer: Singapore, 2022; Volume 29, pp. 239–265, ISBN 978-981-19-3325-7.

30. Nukusheva, A.; Ilyassova, G.; Rustembekova, D.; Zhamiyeva, R.; Arenova, L. Global Warming Problem Faced by the International Community: International Legal Aspect. *Int. Environ. Agreements Politics Law Econ.* **2021**, *21*, 219–233. [CrossRef]
31. Xiong, J.; Yang, Y. Climate Change and Hydrological Extremes. *Curr. Clim. Chang. Rep.* **2024**, *11*, 1. [CrossRef]
32. Sulistiawati, L.Y. Climate Change Related Litigation in Indonesia. *Commun. Earth Environ.* **2024**, *5*, 522. [CrossRef]
33. Malik, I.H.; Ford, J.D. Monitoring Climate Change Vulnerability in the Himalayas. *Ambio* **2025**, *54*, 1–19. [CrossRef]
34. Du Plessis, A. Primary Water Quality Challenges, Contaminants and the World's Dirtiest Places. In *Water as an Inescapable Risk*; Springer Water; Springer International Publishing: Cham, Switzerland, 2019; pp. 79–114, ISBN 978-3-030-03185-5.
35. Singh, N.; Poonia, T.; Siwal, S.S.; Srivastav, A.L.; Sharma, H.K.; Mittal, S.K. Challenges of Water Contamination in Urban Areas. In *Current Directions in Water Scarcity Research*; Elsevier: Amsterdam, The Netherlands, 2022; Volume 6, pp. 173–202, ISBN 978-0-323-91838-1.
36. McDonald, T.L. Review of Environmental Monitoring Methods: Survey Designs. *Environ. Monit. Assess.* **2003**, *85*, 277–292. [CrossRef] [PubMed]
37. Lovett, G.M.; Burns, D.A.; Driscoll, C.T.; Jenkins, J.C.; Mitchell, M.J.; Rustad, L.; Shanley, J.B.; Likens, G.E.; Haeuber, R. Who Needs Environmental Monitoring? *Front. Ecol. Environ.* **2007**, *5*, 253–260. [CrossRef]
38. Kumar, A.; Kim, H.; Hancke, G.P. Environmental Monitoring Systems: A Review. *IEEE Sens. J.* **2013**, *13*, 1329–1339. [CrossRef]
39. Namieśnik, J. Trends in Environmental Analytics and Monitoring. *Crit. Rev. Anal. Chem.* **2000**, *30*, 221–269. [CrossRef]
40. Bourgeois, W.; Romain, A.-C.; Nicolas, J.; Stuetz, R.M. The Use of Sensor Arrays for Environmental Monitoring: Interests and Limitations. *J. Environ. Monit.* **2003**, *5*, 852. [CrossRef] [PubMed]
41. Brammer, J.R.; Brunet, N.D.; Burton, A.C.; Cuerrier, A.; Danielsen, F.; Dewan, K.; Herrmann, T.M.; Jackson, M.V.; Kennett, R.; Larocque, G.; et al. The Role of Digital Data Entry in Participatory Environmental Monitoring. *Conserv. Biol.* **2016**, *30*, 1277–1287. [CrossRef] [PubMed]
42. Trevathan, J.; Johnstone, R. Smart Environmental Monitoring and Assessment Technologies (SEMAT)—A New Paradigm for Low-Cost, Remote Aquatic Environmental Monitoring. *Sensors* **2018**, *18*, 2248. [CrossRef] [PubMed]
43. Abraham, S.; Beard, J.; Manijacob, R. Remote Environmental Monitoring Using Internet of Things (IoT). In Proceedings of the 2017 IEEE Global Humanitarian Technology Conference (GHTC), San Jose, CA, USA, 19–22 October 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–6.
44. Ullo, S.L.; Sinha, G.R. Advances in Smart Environment Monitoring Systems Using IoT and Sensors. *Sensors* **2020**, *20*, 3113. [CrossRef] [PubMed]
45. Chamara, N.; Islam, M.D.; Bai, G.; Shi, Y.; Ge, Y. Ag-IoT for Crop and Environment Monitoring: Past, Present, and Future. *Agric. Syst.* **2022**, *203*, 103497. [CrossRef]
46. Fang, S.; Xu, L.D.; Zhu, Y.; Ahati, J.; Pei, H.; Yan, J.; Liu, Z. An Integrated System for Regional Environmental Monitoring and Management Based on Internet of Things. *IEEE Trans. Ind. Inform.* **2014**, *10*, 1596–1605. [CrossRef]
47. Kishorebabu, V.; Sravanthi, R. Real Time Monitoring of Environmental Parameters Using IOT. *Wirel. Pers. Commun.* **2020**, *112*, 785–808. [CrossRef]
48. Narayana, T.L.; Venkatesh, C.; Kiran, A.; J, C.B.; Kumar, A.; Khan, S.B.; Almusharraf, A.; Quasim, M.T. Advances in Real Time Smart Monitoring of Environmental Parameters Using IoT and Sensors. *Heliyon* **2024**, *10*, e28195. [CrossRef] [PubMed]
49. Essamlali, I.; Nhaila, H.; El Khaili, M. Advances in Machine Learning and IoT for Water Quality Monitoring: A Comprehensive Review. *Heliyon* **2024**, *10*, e27920. [CrossRef] [PubMed]
50. Forhad, H.M.; Uddin, M.R.; Chakrovorty, R.S.; Ruhul, A.M.; Faruk, H.M.; Kamruzzaman, S.; Sharmin, N.; Jamal, A.S.I.M.; Haque, M.M.-U.; Morshed, A.M. IoT Based Real-Time Water Quality Monitoring System in Water Treatment Plants (WTPs). *Heliyon* **2024**, *10*, e40746. [CrossRef] [PubMed]
51. Hino, M.; Benami, E.; Brooks, N. Machine Learning for Environmental Monitoring. *Nat. Sustain.* **2018**, *1*, 583–588. [CrossRef]
52. Jesus, G.; Casimiro, A.; Oliveira, A. Using Machine Learning for Dependable Outlier Detection in Environmental Monitoring Systems. *ACM Trans. Cyber-Phys. Syst.* **2021**, *5*, 1–30. [CrossRef]
53. Kubečka, J.; Knattrup, Y.; Engsvang, M.; Jensen, A.B.; Ayoubi, D.; Wu, H.; Christiansen, O.; Elm, J. Current and Future Machine Learning Approaches for Modeling Atmospheric Cluster Formation. *Nat. Comput. Sci.* **2023**, *3*, 495–503. [CrossRef] [PubMed]
54. Esenogho, E.; Djouani, K.; Kurien, A.M. Integrating Artificial Intelligence Internet of Things and 5G for Next-Generation Smartgrid: A Survey of Trends Challenges and Prospect. *IEEE Access* **2022**, *10*, 4794–4831. [CrossRef]
55. Cortés, U.; Sànchez-Marrè, M.; Ceccaroni, L.; R-Roda, I.; Poch, M. Artificial Intelligence and Environmental Decision Support Systems. *Appl. Intell.* **2000**, *13*, 77–91. [CrossRef]
56. Haupt, S.E.; Pasini, A.; Marzban, C. (Eds.) *Artificial Intelligence Methods in the Environmental Sciences*; Springer: Dordrecht, The Netherlands, 2009; ISBN 978-1-4020-9117-9.
57. Shuford, J. Interdisciplinary Perspectives: Fusing Artificial Intelligence with Environmental Science for Sustainable Solutions. *J. Artif. Intell. Gen. Sci. (JAIGS)* **2024**, *1*, 106–123. [CrossRef]

58. Himeur, Y.; Rimal, B.; Tiwary, A.; Amira, A. Using Artificial Intelligence and Data Fusion for Environmental Monitoring: A Review and Future Perspectives. *Inf. Fusion* **2022**, *86–87*, 44–75. [CrossRef]
59. Szramowiat-Sala, K. Artificial Intelligence in Environmental Monitoring: Application of Artificial Neural Networks and Machine Learning for Pollution Prevention and Toxicity Measurements. 2023. Available online: https://www.preprints.org/manuscript/202307.1298 (accessed on 23 December 2024).
60. Popescu, S.M.; Mansoor, S.; Wani, O.A.; Kumar, S.S.; Sharma, V.; Sharma, A.; Arya, V.M.; Kirkham, M.B.; Hou, D.; Bolan, N.; et al. Artificial Intelligence and IoT Driven Technologies for Environmental Pollution Monitoring and Management. *Front. Environ. Sci.* **2024**, *12*, 1336088. [CrossRef]
61. Yetilmezsoy, K.; Ozkaya, B.; Cakmakci, M. Artificial Intelligence-Based Prediction Models for Environmental Engineering. *Neural Netw. World* **2011**, *21*, 193–218. [CrossRef]
62. Bagheri, M.; Bazvand, A.; Ehteshami, M. Application of Artificial Intelligence for the Management of Landfill Leachate Penetration into Groundwater, and Assessment of Its Environmental Impacts. *J. Clean. Prod.* **2017**, *149*, 784–796. [CrossRef]
63. Maganathan, T.; Senthilkumar, S.; Balakrishnan, V. Machine Learning and Data Analytics for Environmental Science: A Review, Prospects and Challenges. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *955*, 012107. [CrossRef]
64. Kosovic, I.N.; Mastelic, T.; Ivankovic, D. Using Artificial Intelligence on Environmental Data from Internet of Things for Estimating Solar Radiation: Comprehensive Analysis. *J. Clean. Prod.* **2020**, *266*, 121489. [CrossRef]
65. Suguna, S.K.; Dhivya, M.; Paiva, S. (Eds.) *Artificial Intelligence (AI): Recent Trends and Applications*; CRC Press: Boca Raton, FL, USA; Taylor & Francis Group: London, UK; New York, NY, USA, 2021; ISBN 978-1-003-00562-9.
66. Wankhede, V.A.; Agrawal, R.; Kumar, A.; Luthra, S.; Pamucar, D.; Stević, Ž. Artificial Intelligence an Enabler for Sustainable Engineering Decision-Making in Uncertain Environment: A Review and Future Propositions. *J. Glob. Oper. Strateg. Sourc.* **2024**, *17*, 384–401. [CrossRef]
67. Wooldridge, M.; Jennings, N.R. Intelligent Agents: Theory and Practice. *Knowl. Eng. Rev.* **1995**, *10*, 115–152. [CrossRef]
68. Kabanza, F.; Barbeau, M.; St-Denis, R. Planning Control Rules for Reactive Agents. *Artif. Intell.* **1997**, *95*, 67–113. [CrossRef]
69. Rebera, A.P. Reactive Attitudes and AI-Agents–Making Sense of Responsibility and Control Gaps. *Philos. Technol.* **2024**, *37*, 126. [CrossRef]
70. Hayes-Roth, B. Integrating Real-Time AI Techniques in Adaptive Intelligent Agents. In *Artificial Intelligence in Real-Time Control 1994*; Elsevier: Oxford, UK, 1995; pp. 1–11, ISBN 978-0-08-042236-7.
71. Mounce, S.R.; Boxall, J.B.; Machell, J. Development and Verification of an Online Artificial Intelligence System for Detection of Bursts and Other Abnormal Flows. *J. Water Resour. Plan. Manag.* **2010**, *136*, 309–318. [CrossRef]
72. Ontañón, S.; Plaza, E. Learning and Joint Deliberation through Argumentation in Multiagent Systems. In Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems, Honolulu, HI, USA, 14–18 May 2007; ACM: New York, NY, USA, 2007; pp. 1–8.
73. Corchado, J.M.; Laza, R. Constructing Deliberative Agents with Case-Based Reasoning Technology. *Int. J. Intell. Syst.* **2003**, *18*, 1227–1241. [CrossRef]
74. Boddy, M.; Dean, T.L. Deliberation Scheduling for Problem Solving in Time-Constrained Environments. *Artif. Intell.* **1994**, *67*, 245–285. [CrossRef]
75. Singh, M.; Arora, V.; Kulshreshta, K. AI and the Environment: Innovative Approaches to Climate Change. In *Practice, Progress, and Proficiency in Sustainability*; Singh, B., Kaunert, C., Vig, K., Dutta, S., Eds.; IGI Global: Hershey, PA, USA, 2024; pp. 1–22, ISBN 979-8-3693-6336-2.
76. Zhao, T.; Wang, S.; Ouyang, C.; Chen, M.; Liu, C.; Zhang, J.; Yu, L.; Wang, F.; Xie, Y.; Li, J.; et al. Artificial Intelligence for Geoscience: Progress, Challenges, and Perspectives. *Innovation* **2024**, *5*, 100691. [CrossRef]
77. Guessoum, Z. A Hybrid Agent Model: A Reactive and Cognitive Behavior. In Proceedings of the Third International Symposium on Autonomous Decentralized Systems (ISADS 97), Berlin, Germany, 9–11 April 1997; IEEE Computer Society Press: Los Alamitos, CA, USA, 1997; pp. 25–32.
78. Palanca, J.; Rincon, J.A.; Carrascosa, C.; Julian, V.J.; Terrasa, A. Flexible Agent Architecture: Mixing Reactive and Deliberative Behaviors in SPADE. *Electronics* **2023**, *12*, 659. [CrossRef]
79. De Paola, A.; Ferraro, P.; Lo Re, G.; Morana, M.; Ortolani, M. A Fog-Based Hybrid Intelligent System for Energy Saving in Smart Buildings. *J. Ambient. Intell. Humaniz. Comput.* **2020**, *11*, 2793–2807. [CrossRef]
80. Janssen, M.; De Vries, B. The Battle of Perspectives: A Multi-Agent Model with Adaptive Responses to Climate Change. *Ecol. Econ.* **1998**, *26*, 43–65. [CrossRef]
81. Figueiredo, J.; Botto, M.A.; Rijo, M. SCADA System with Predictive Controller Applied to Irrigation Canals. *Control Eng. Pract.* **2013**, *21*, 870–886. [CrossRef]
82. Ayala-Cabrera, D.; Herrera, M.; Izquierdo, J.; Pérez-García, R. GPR Data Analysis Using Multi-Agent and Clustering Approaches: A Tool for Technical Management of Water Supply Systems. *Digit. Signal Process.* **2014**, *27*, 140–149. [CrossRef]

83. Asha, P.; Natrayan, L.; Geetha, B.T.; Beulah, J.R.; Sumathy, R.; Varalakshmi, G.; Neelakandan, S. IoT Enabled Environmental Toxicology for Air Pollution Monitoring Using AI Techniques. *Environ. Res.* **2022**, *205*, 112574. [CrossRef] [PubMed]
84. Nti, E.K.; Cobbina, S.J.; Attafuah, E.E.; Senanu, L.D.; Amenyeku, G.; Gyan, M.A.; Forson, D.; Safo, A.-R. Water Pollution Control and Revitalization Using Advanced Technologies: Uncovering Artificial Intelligence Options towards Environmental Health Protection, Sustainability and Water Security. *Heliyon* **2023**, *9*, e18170. [CrossRef] [PubMed]
85. Castelfranchi, C.; Dignum, F.; Jonker, C.M.; Treur, J. Deliberative Normative Agents: Principles and Architecture. In *Intelligent Agents VI. Agent Theories, Architectures, and Languages*; Jennings, N.R., Lespérance, Y., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2000; Volume 1757, pp. 364–378. [CrossRef]
86. Hare, M.; Deadman, P. Further Towards a Taxonomy of Agent-Based Simulation Models in Environmental Management. *Math. Comput. Simul.* **2004**, *64*, 25–40. [CrossRef]
87. Sánchez-Fibla, M.; Moulin-Frier, C.; Solé, R. Cooperative Control of Environmental Extremes by Artificial Intelligent Agents. *J. R. Soc. Interface* **2024**, *21*, 20240344. [CrossRef] [PubMed]
88. Bezborodova, O.E.; Bodin, O.N.; Paschenko, D.V. Improving Data Collection and Processing Efficiency by Using Hierarchies of Intelligent Agents in Multi-Agent Systems. *IOP Conf. Ser. Mater. Sci. Eng.* **2021**, *1061*, 012009. [CrossRef]
89. Grzonka, D.; Jakóbik, A.; Kołodziej, J.; Pllana, S. Using a Multi-Agent System and Artificial Intelligence for Monitoring and Improving the Cloud Performance and Security. *Future Gener. Comput. Syst.* **2018**, *86*, 1106–1117. [CrossRef]
90. Zhang, X.; Shu, K.; Rajkumar, S.; Sivakumar, V. Research on Deep Integration of Application of Artificial Intelligence in Environmental Monitoring System and Real Economy. *Environ. Impact Assess. Rev.* **2021**, *86*, 106499. [CrossRef]
91. Luzolo, P.H.; Elrawashdeh, Z.; Tchappi, I.; Galland, S.; Outay, F. Combining Multi-Agent Systems and Artificial Intelligence of Things: Technical Challenges and Gains. *Internet Things* **2024**, *28*, 101364. [CrossRef]
92. Vald, G.; Sermet, M.; Mount, J.; Shrestha, S.; Samuel, D.J.; Cwiertny, D.; Demir, I. Integrating Conversational AI Agents for Enhanced Water Quality Analytics: Development of a Novel Data Expert System. 2024, *in press*.
93. Hart, J.K.; Martinez, K. Environmental Sensor Networks: A Revolution in the Earth System Science? *Earth-Sci. Rev.* **2006**, *78*, 177–191. [CrossRef]
94. El Khediri, S.; Benfradj, A.; Thaljaoui, A.; Moulahi, T.; Ullah Khan, R.; Alabdulatif, A.; Lorenz, P. Integration of Artificial Intelligence (AI) with Sensor Networks: Trends, Challenges, and Future Directions. *J. King Saud Univ.-Comput. Inf. Sci.* **2024**, *36*, 101892. [CrossRef]
95. Janga, B.; Asamani, G.; Sun, Z.; Cristea, N. A Review of Practical AI for Remote Sensing in Earth Sciences. *Remote Sens.* **2023**, *15*, 4112. [CrossRef]
96. Solomou, S.; Sengupta, U. Simulating Complex Urban Behaviours with AI: Incorporating Improved Intelligent Agents in Urban Simulation Models. *Urban Plan.* **2024**, *10*, 8561. [CrossRef]
97. Bouziane, S.E.; Khadir, M.T.; Dugdale, J. A Collaborative Predictive Multi-Agent System for Forecasting Carbon Emissions Related to Energy Consumption. *Multiagent Grid Syst.* **2021**, *17*, 39–58. [CrossRef]
98. Jones, J.; Harris, E.; Febriansah, Y.; Adiwijaya, A.; Nuril Hikam, I. AI for Sustainable Development: Applications in Natural Resource Management, Agriculture, and Waste Management. *Int. Trans. Artif. Intell. (ITALIC)* **2024**, *2*, 143–149. [CrossRef]
99. Kumar, S.; Verma, A.K.; Mirza, A. Artificial Intelligence and Climate Change Mitigation. In *Digital Transformation, Artificial Intelligence and Society*; Frontiers of Artificial Intelligence, Ethics and Multidisciplinary Applications; Springer: Singapore, 2024; pp. 147–160, ISBN 978-981-97-5655-1.
100. Amiri, Z.; Heidari, A.; Navimipour, N.J. Comprehensive Survey of Artificial Intelligence Techniques and Strategies for Climate Change Mitigation. *Energy* **2024**, *308*, 132827. [CrossRef]
101. Alawi, O.A.; Kamar, H.M.; Alsuwaiyan, A.; Yaseen, Z.M. Temporal Trends and Predictive Modeling of Air Pollutants in Delhi: A Comparative Study of Artificial Intelligence Models. *Sci. Rep.* **2024**, *14*, 30957. [CrossRef] [PubMed]
102. Reza, S.A.; Chowdhury, M.S.R.; Hossain, S.; Hasanuzzaman, M.; Shawon, R.E.R.; Chowdhury, B.R.; Rana, M.S. Global Plastic Waste Management: Analyzing Trends, Economic and Social Implications, and Predictive Modeling Using Artificial Intelligence. *J. Environ. Agric. Stud.* **2024**, *5*, 42–58. [CrossRef]
103. Lee, C.C.; Barnes, B.B.; Sheridan, S.C.; Smith, E.T.; Hu, C.; Pirhalla, D.E.; Ransibrahmanakul, V.; Adams, R. Using Machine Learning to Model and Predict Water Clarity in the Great Lakes. *J. Great Lakes Res.* **2020**, *46*, 1501–1510. [CrossRef]
104. Raihan, A. Artificial Intelligence and Machine Learning Applications in Forest Management and Biodiversity Conservation. *Nat. Resour. Conserv. Res.* **2023**, *6*, 3825. [CrossRef]
105. Ullah, F.; Saqib, S.; Xiong, Y.-C. Integrating Artificial Intelligence in Biodiversity Conservation: Bridging Classical and Modern Approaches. *Biodivers. Conserv.* **2024**. [CrossRef]
106. Ayoola, V.B.; Idoko, I.P.; Eromonsei, S.O.; Afolabi, O.; Apampa, A.R.; Oyebanji, O.S. The Role of Big Data and AI in Enhancing Biodiversity Conservation and Resource Management in the USA. *World J. Adv. Res. Rev.* **2024**, *23*, 1851–1873. [CrossRef]
107. Silvestro, D.; Goria, S.; Sterner, T.; Antonelli, A. Improving Biodiversity Protection through Artificial Intelligence. *Nat. Sustain.* **2022**, *5*, 415–424. [CrossRef]

108. Kamrowska-Załuska, D. Impact of AI-Based Tools and Urban Big Data Analytics on the Design and Planning of Cities. *Land* **2021**, *10*, 1209. [CrossRef]
109. Yigitcanlar, T.; Kankanamge, N.; Regona, M.; Ruiz Maldonado, A.; Rowan, B.; Ryu, A.; Desouza, K.C.; Corchado, J.M.; Mehmood, R.; Li, R.Y.M. Artificial Intelligence Technologies and Related Urban Planning and Development Concepts: How Are They Perceived and Utilized in Australia? *J. Open Innov. Technol. Mark. Complex.* **2020**, *6*, 187. [CrossRef]
110. Jha, A.K.; Ghimire, A.; Thapa, S.; Jha, A.M.; Raj, R. A Review of AI for Urban Planning: Towards Building Sustainable Smart Cities. In Proceedings of the 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 20–22 January 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 937–944.
111. Sanchez, T.W.; Shumway, H.; Gordner, T.; Lim, T. The Prospects of Artificial Intelligence in Urban Planning. *Int. J. Urban Sci.* **2023**, *27*, 179–194. [CrossRef]
112. Chen, C.-Y.; Chai, K.K.; Lau, E. AI-Assisted Approach for Building Energy and Carbon Footprint Modeling. *Energy AI* **2021**, *5*, 100091. [CrossRef]
113. Aras, S.; Hanifi Van, M. An Interpretable Forecasting Framework for Energy Consumption and $CO_2$ Emissions. *Appl. Energy* **2022**, *328*, 120163. [CrossRef]
114. Nazir, T.; Nagra, H.F.; Bhatti, M.H.; Shahid, R.N.; Shaukat, N.; Tariq, N.U.H. Predictive Analytics for Environmental Impact: Deep Learning and AI Ensemble Strategies in Forecasting $CO_2$ Emissions. 2024. Available online: https://www.researchsquare.com/article/rs-4081410/v1 (accessed on 23 December 2024).
115. Meng, Y.; Noman, H. Predicting $CO_2$ Emission Footprint Using AI through Machine Learning. *Atmosphere* **2022**, *13*, 1871. [CrossRef]
116. Gaur, L.; Afaq, A.; Arora, G.K.; Khan, N. Artificial Intelligence for Carbon Emissions Using System of Systems Theory. *Ecol. Inform.* **2023**, *76*, 102165. [CrossRef]
117. Brady, M. Artificial Intelligence and Robotics. *Artif. Intell.* **1985**, *26*, 79–121. [CrossRef]
118. Rajan, K.; Saffiotti, A. Towards a Science of Integrated AI and Robotics. *Artif. Intell.* **2017**, *247*, 1–9. [CrossRef]
119. Moglia, A.; Georgiou, K.; Georgiou, E.; Satava, R.M.; Cuschieri, A. A Systematic Review on Artificial Intelligence in Robot-Assisted Surgery. *Int. J. Surg.* **2021**, *95*, 106151. [CrossRef]
120. Giordano, G.; Murali Babu, S.P.; Mazzolai, B. Soft Robotics towards Sustainable Development Goals and Climate Actions. *Front. Robot. AI* **2023**, *10*, 1116005. [CrossRef]
121. Putta, P.; Mills, E.; Garg, N.; Motwani, S.; Finn, C.; Garg, D.; Rafailov, R. Agent Q: Advanced Reasoning and Learning for Autonomous AI Agents. *arXiv* **2024**, arXiv:2408.07199.
122. Maes, P. Modeling Adaptive Autonomous Agents. *Artif. Life* **1993**, *1*, 135–162. [CrossRef]
123. Hauptman, A.I.; Schelble, B.G.; McNeese, N.J.; Madathil, K.C. Adapt and Overcome: Perceptions of Adaptive Autonomous Agents for Human-AI Teaming. *Comput. Hum. Behav.* **2023**, *138*, 107451. [CrossRef]
124. Totschnig, W. Fully Autonomous AI. *Sci. Eng. Ethics* **2020**, *26*, 2473–2485. [CrossRef] [PubMed]
125. Candrian, C.; Scherer, A. Rise of the Machines: Delegating Decisions to Autonomous AI. *Comput. Hum. Behav.* **2022**, *134*, 107308. [CrossRef]
126. Hirzle, T.; Müller, F.; Draxler, F.; Schmitz, M.; Knierim, P.; Hornbæk, K. When XR and AI Meet–A Scoping Review on Extended Reality and Artificial Intelligence. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, Hamburg, Germany, 19–24 April 2023; ACM: New York, NY, USA, 2023; pp. 1–45.
127. Reiners, D.; Davahli, M.R.; Karwowski, W.; Cruz-Neira, C. The Combination of Artificial Intelligence and Extended Reality: A Systematic Review. *Front. Virtual Real.* **2021**, *2*, 721933. [CrossRef]
128. Pei, J.; Viola, I.; Huang, H.; Wang, J.; Ahsan, M.; Ye, F.; Yiming, J.; Sai, Y.; Wang, D.; Chen, Z.; et al. Autonomous Workflow for Multimodal Fine-Grained Training Assistants Towards Mixed Reality. *arXiv* **2024**, arXiv:2405.13034.
129. Stanney, K.M.; Archer, J.; Skinner, A.; Horner, C.; Hughes, C.; Brawand, N.P.; Martin, E.; Sanchez, S.; Moralez, L.; Fidopiastis, C.M.; et al. Performance Gains from Adaptive eXtended Reality Training Fueled by Artificial Intelligence. *J. Def. Model. Simul. Appl. Methodol. Technol.* **2022**, *19*, 195–218. [CrossRef]
130. Serrano, M.; Dang, H.N.; Nguyen, H.M.Q. Recent Advances on Artificial Intelligence and Internet of Things Convergence for Human-Centric Applications: Internet of Things Science. In Proceedings of the 8th International Conference on the Internet of Things (IoT), Santa Barbara, CA, USA, 15–18 October 2018; ACM: New York, NY, USA, 2018; pp. 1–5.
131. Shi, F.; Ning, H.; Huangfu, W.; Zhang, F.; Wei, D.; Hong, T.; Daneshmand, M. Recent Progress on the Convergence of the Internet of Things and Artificial Intelligence. *IEEE Netw.* **2020**, *34*, 8–15. [CrossRef]
132. Mastorakis, G.; Mavromoustakis, C.X.; Batalla, J.M.; Pallis, E. (Eds.) *Convergence of Artificial Intelligence and the Internet of Things*; Internet of Things Series; Springer International Publishing: Cham, Switzerland, 2020; ISBN 978-3-030-44906-3.
133. Bibri, S.E.; Alexandre, A.; Sharifi, A.; Krogstie, J. Environmentally Sustainable Smart Cities and Their Converging AI, IoT, and Big Data Technologies and Solutions: An Integrated Approach to an Extensive Literature Review. *Energy Inform.* **2023**, *6*, 9. [CrossRef] [PubMed]

134. Khan, S.A.; Kalifullah, A.H.; Ibragimova, K.; Singh, A.K.; Muniyandy, E.; Rachapudi, V. Integrating AI and IoT in Advanced Optical Systems for Sustainable Energy and Environment Monitoring. *Int. J. Adv. Comput. Sci. Appl.* **2024**, *15*. [CrossRef]
135. Yu, Q.; Xiong, F.; Wang, Y. Integration of Wireless Sensor Network and IoT for Smart Environment Monitoring System. *J. Interconnect. Netw.* **2022**, *22*, 2143010. [CrossRef]
136. Bin Mofidul, R.; Alam, M.M.; Rahman, M.H.; Jang, Y.M. Real-Time Energy Data Acquisition, Anomaly Detection, and Monitoring System: Implementation of a Secured, Robust, and Integrated Global IIoT Infrastructure with Edge and Cloud AI. *Sensors* **2022**, *22*, 8980. [CrossRef] [PubMed]
137. Sharma, A.; Singh, K.J.; Kapoor, D.S.; Thakur, K.; Mahajan, S. The Role of IoT in Environmental Sustainability: Advancements and Applications for Smart Cities. In *Mobile Crowdsensing and Remote Sensing in Smart Cities*; Kerrache, C.A., Sahraoui, Y., Calafate, C.T., Vegni, A.M., Eds.; Internet of Things; Springer: Cham, Switzerland, 2025; pp. 21–39, ISBN 978-3-031-72731-3.
138. Selvam, A.P.; Al-Humairi, S.N.S. *The Impact of IoT and Sensor Integration on Real-Time Weather Monitoring Systems: A Systematic Review*; Springer Science and Business Media LLC: Berlin, Germany, 2023.
139. Rahu, M.A.; Karim, S.; Ali, S.M.; Jatoi, G.M.; Sohu, N.D. Integration of Wireless Sensor Networks, Internet of Things, Artificial Intelligence, and Deep Learning in Smart Agriculture: A Comprehensive Survey. *J. Innov. Intell. Comput. Emerg. Technol. (JIICET)* **2024**, *1*, 8–18.
140. Alam, M.J.B.; Manzano, L.S.; Debnath, R.; Ahmed, A.A. Monitoring Slope Movement and Soil Hydrologic Behavior Using IoT and AI Technologies: A Systematic Review. *Hydrology* **2024**, *11*, 111. [CrossRef]
141. Ranjan, R.; Rana, O.; Nepal, S.; Yousif, M.; James, P.; Wen, Z.; Barr, S.; Watson, P.; Jayaraman, P.P.; Georgakopoulos, D.; et al. The Next Grand Challenges: Integrating the Internet of Things and Data Science. *IEEE Cloud Comput.* **2018**, *5*, 12–26. [CrossRef]
142. Meer, T.; Gupta, R.; Ailyn, D. Smart Sensor Integration: IoT Sensors for Real-Time Monitoring of Environmental Conditions. 2024. Available online: https://www.researchgate.net/publication/384402035_Smart_Sensor_Integration_IoT_sensors_for_real-time_monitoring_of_environmental_conditions (accessed on 27 December 2024).
143. Bibri, S.E.; Krogstie, J.; Kaboli, A.; Alahi, A. Smarter Eco-Cities and Their Leading-Edge Artificial Intelligence of Things Solutions for Environmental Sustainability: A Comprehensive Systematic Review. *Environ. Sci. Ecotechnol.* **2024**, *19*, 100330. [CrossRef] [PubMed]
144. Manavalan, M. Intersection of Artificial Intelligence, Machine Learning, and Internet of Things–An Economic Overview. *Glob. Discl. Econ. Bus.* **2020**, *9*, 119–128. [CrossRef]
145. Kalusivalingam, A.K.; Sharma, A.; Patel, N.; Singh, V. Enhancing Smart City Development with AI: Leveraging Machine Learning Algorithms and IoT-Driven Data Analytics. *Int. J. AI ML* **2021**, *2*.
146. Mazhar, T.; Irfan, H.M.; Haq, I.; Ullah, I.; Ashraf, M.; Shloul, T.A.; Ghadi, Y.Y.; Imran; Elkamchouchi, D.H. Analysis of Challenges and Solutions of IoT in Smart Grids Using AI and Machine Learning Techniques: A Review. *Electronics* **2023**, *12*, 242. [CrossRef]
147. Ali, S.I.; Abdulqader, D.M.; Ahmed, O.M.; Ismael, H.R.; Hasan, S.; Ahmed, L.H. Consideration of Web Technology and Cloud Computing Inspiration for AI and IoT Role in Sustainable Decision-Making for Enterprise Systems. *J. Inf. Technol. Inform.* **2024**, *3*, 4.
148. Protopsaltis, A.; Sarigiannidis, P.; Margounakis, D.; Lytos, A. Data Visualization in Internet of Things: Tools, Methodologies, and Challenges. In Proceedings of the 15th International Conference on Availability, Reliability and Security, Virtual Event, Ireland, 25 August 2020; ACM: New York, NY, USA, 2020; pp. 1–11.
149. Salamkar, M.A. Data Visualization: AI-Enhanced Visualization Tools to Better Interpret Complex Data Patterns. *J. Bioinform. Artif. Intell.* **2024**, *4*, 204–226. [CrossRef]
150. Singh, J. Sensor-Based Personal Data Collection in the Digital Age: Exploring Privacy Implications, AI-Driven Analytics, and Security Challenges in IoT and Wearable Devices. *Distrib. Learn. Broad Appl. Sci. Res.* **2019**, *5*, 785–809.
151. Kanungo, S. Revolutionizing Data Processing: Advanced Cloud Computing and AI Synergy for IoT Innovation. *Int. Res. J. Mod. Eng. Technol. Sci.* **2020**, *2*, 1032–1040.
152. Tawakuli, A.; Havers, B.; Gulisano, V.; Kaiser, D.; Engel, T. Survey: Time-Series Data Preprocessing: A Survey and an Empirical Analysis. *J. Eng. Res.* 2024, *in press*.
153. Dhawas, P.; Ramteke, M.A.; Thakur, A.; Polshetwar, P.V.; Salunkhe, R.V.; Bhagat, D. Big Data Analysis Techniques: Data Preprocessing Techniques, Data Mining Techniques, Machine Learning Algorithm, Visualization. In *Big Data Analytics Techniques for Market Intelligence*; IGI Global: Hershey, PA, USA, 2024; pp. 183–208.
154. Pattyam, S.P. AI-Driven Data Science for Environmental Monitoring: Techniques for Data Collection, Analysis, and Predictive Modeling. *Aust. J. Mach. Learn. Res. Appl.* **2021**, *1*, 132–169.
155. Panduman, Y.Y.F.; Funabiki, N.; Fajrianti, E.D.; Fang, S.; Sukaridhoto, S. A Survey of AI Techniques in IoT Applications with Use Case Investigations in the Smart Environmental Monitoring and Analytics in Real-Time IoT Platform. *Information* **2024**, *15*, 153. [CrossRef]
156. Ramadan, M.N.A.; Ali, M.A.H.; Khoo, S.Y.; Alkhedher, M.; Alherbawi, M. Real-Time IoT-Powered AI System for Monitoring and Forecasting of Air Pollution in Industrial Environment. *Ecotoxicol. Environ. Saf.* **2024**, *283*, 116856. [CrossRef]

157. Villegas-Ch, W.; García-Ortiz, J.; Sánchez-Viteri, S. Towards Intelligent Monitoring in IoT: AI Applications for Real-Time Analysis and Prediction. *IEEE Access* 2024, *in press*.
158. Habeeb, R.A.A.; Nasaruddin, F.; Gani, A.; Hashem, I.A.T.; Ahmed, E.; Imran, M. Real-Time Big Data Processing for Anomaly Detection: A Survey. *Int. J. Inf. Manag.* **2019**, *45*, 289–307. [CrossRef]
159. Trilles, S.; Belmonte, Ò.; Schade, S.; Huerta, J. A Domain-Independent Methodology to Analyze IoT Data Streams in Real-Time. A Proof of Concept Implementation for Anomaly Detection from Environmental Data. *Int. J. Digit. Earth* **2017**, *10*, 103–120. [CrossRef]
160. Srivastava, A.; Maity, R. Assessing the Potential of AI–ML in Urban Climate Change Adaptation and Sustainable Development. *Sustainability* **2023**, *15*, 16461. [CrossRef]
161. 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]
162. Bala Dhandayuthapani, V. AI, IoT, and Smart Technologies for Environmental Resilience and Sustainability–Comprehensive Review. *Int. J. Inf. Eng. Electron. Bus.* **2024**, *16*, 75–84.
163. Muhammad, A. Managing River Basins with Thinking Machines. In Proceedings of the 2016 IEEE Conference on Norbert Wiener in the 21st Century (21CW), Melbourne, Australia, 13–15 July 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1–6.
164. Moreno, C.; Aquino, R.; Ibarreche, J.; Pérez, I.; Castellanos, E.; Álvarez, E.; Rentería, R.; Anguiano, L.; Edwards, A.; Lepper, P. RiverCore: IoT Device for River Water Level Monitoring over Cellular Communications. *Sensors* **2019**, *19*, 127. [CrossRef] [PubMed]
165. Zakaria, M.I.; Jabbar, W.A.; Sulaiman, N. Development of a Smart Sensing Unit for LoRaWAN-Based IoT Flood Monitoring and Warning System in Catchment Areas. *Internet Things Cyber-Phys. Syst.* **2023**, *3*, 249–261. [CrossRef]
166. Wei, J.; Liu, S.; Li, Z.; Liu, C.; Qin, K.; Liu, X.; Pinker, R.T.; Dickerson, R.R.; Lin, J.; Boersma, K.F.; et al. Ground-Level $NO_2$ Surveillance from Space Across China for High Resolution Using Interpretable Spatiotemporally Weighted Artificial Intelligence. *Environ. Sci. Technol.* **2022**, *56*, 9988–9998. [CrossRef]
167. Arat, M. Detection of Anomalous Nitrogen Dioxide ($NO_2$) Concentration of a District in Ankara: A Reconstruction-Based Approach. *J. Polytech.-Politek Derg.* 2024, *in press*.
168. Divya, A.; Kavithanjali, T.; Dharshini, P. IoT Enabled Forest Fire Detection and Early Warning System. In Proceedings of the 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), Puducherry, India, 29–30 March 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–5. [CrossRef]
169. Chan, C.C.; Alvi, S.A.; Zhou, X.; Durrani, S.; Wilson, N.; Yebra, M. A Survey on IoT Ground Sensing Systems for Early Wildfire Detection: Technologies, Challenges and Opportunities. *IEEE Access* 2024, *in press*.
170. Bhardwaj, N.; Joshi, P. A MATTER-Enabled IoT Framework for Enhanced Fire Detection and Real-Time Decision-Making. *SN Comput. Sci.* **2024**, *5*, 1088. [CrossRef]
171. Kanakaraja, P.; Sundar, P.S.; Vaishnavi, N.; Reddy, S.G.K.; Manikanta, G.S. IoT Enabled Advanced Forest Fire Detecting and Monitoring on Ubidots Platform. *Mater. Today Proc.* **2021**, *46*, 3907–3914. [CrossRef]
172. Manoj, M.; Dhilip Kumar, V.; Arif, M.; Bulai, E.-R.; Bulai, P.; Geman, O. State of the Art Techniques for Water Quality Monitoring Systems for Fish Ponds Using IoT and Underwater Sensors: A Review. *Sensors* **2022**, *22*, 2088. [CrossRef] [PubMed]
173. Zainurin, S.N.; Wan Ismail, W.Z.; Mahamud, S.N.I.; Ismail, I.; Jamaludin, J.; Ariffin, K.N.Z.; Wan Ahmad Kamil, W.M. Advancements in Monitoring Water Quality Based on Various Sensing Methods: A Systematic Review. *Int. J. Environ. Res. Public Health* **2022**, *19*, 14080. [CrossRef]
174. Chowdury, M.S.U.; Emran, T.B.; Ghosh, S.; Pathak, A.; Alam, M.M.; Absar, N.; Andersson, K.; Hossain, M.S. IoT Based Real-Time River Water Quality Monitoring System. *Procedia Comput. Sci.* **2019**, *155*, 161–168. [CrossRef]
175. Prapti, D.R.; Mohamed Shariff, A.R.; Che Man, H.; Ramli, N.M.; Perumal, T.; Shariff, M. Internet of Things (IoT)-Based Aquaculture: An Overview of IoT Application on Water Quality Monitoring. *Rev. Aquac.* **2022**, *14*, 979–992. [CrossRef]
176. Islam, M.M.; Kashem, M.A.; Alyami, S.A.; Moni, M.A. Monitoring Water Quality Metrics of Ponds with IoT Sensors and Machine Learning to Predict Fish Species Survival. *Microprocess. Microsyst.* **2023**, *102*, 104930. [CrossRef]
177. Avancini, D.B.; Rodrigues, J.J.P.C.; Rabêlo, R.A.L.; Das, A.K.; Kozlov, S.; Solic, P. A New IoT-based Smart Energy Meter for Smart Grids. *Int. J. Energy Res.* **2021**, *45*, 189–202. [CrossRef]
178. Mohamed, A.; Mohammed, O. Real-Time Energy Management Scheme for Hybrid Renewable Energy Systems in Smart Grid Applications. *Electr. Power Syst. Res.* **2013**, *96*, 133–143. [CrossRef]
179. Morello, R.; De Capua, C.; Fulco, G.; Mukhopadhyay, S.C. A Smart Power Meter to Monitor Energy Flow in Smart Grids: The Role of Advanced Sensing and IoT in the Electric Grid of the Future. *IEEE Sens. J.* **2017**, *17*, 7828–7837. [CrossRef]
180. Konya, A.; Nematzadeh, P. Recent Applications of AI to Environmental Disciplines: A Review. *Sci. Total. Environ.* **2024**, *906*, 167705. [CrossRef] [PubMed]

181. Santos, M.R.; Carvalho, L.C. AI-Driven Participatory Environmental Management: Innovations, Applications, and Future Prospects. *J. Environ. Manag.* **2025**, *373*, 123864. [CrossRef] [PubMed]
182. Chen, Y. IoT, Cloud, Big Data and AI in Interdisciplinary Domains. *Simul. Model. Pract. Theory* **2020**, *102*, 102070. [CrossRef]
183. Hamza, M.A.; Shaiba, H.; Marzouk, R.; Alhindi, A.; Asiri, M.M.; Yaseen, I.; Motwakel, A.; Rizwanullah, M. Big Data Analytics with Artificial Intelligence Enabled Environmental Air Pollution Monitoring Framework. *Comput. Mater. Contin.* **2022**, *73*, 3235–3250.
184. Kumar, A.; Madaan, G.; Sharma, P.; Kumar, A. Application of Disruptive Technologies on Environmental Health: An Overview of Artificial Intelligence, Blockchain and Internet of Things. *Asia Pac. J. Health Manag.* **2021**, *16*, 251–259. [CrossRef]
185. Basu, M.; Nath, M.D. A Study of the Role of Artificial Intelligence in Monitoring Environmental and Health Issues in the Post-COVID-19 Pandemic Era for Sustainable Living. In *Artificial Intelligence for Multimedia Information Processing*; CRC Press: Boca Raton, FL, USA, 2024; pp. 132–166.
186. Kumari, N.; Pandey, S. Application of Artificial Intelligence in Environmental Sustainability and Climate Change. In *Visualization Techniques for Climate Change with Machine Learning and Artificial Intelligence*; Elsevier: Amsterdam, The Netherlands, 2023; pp. 293–316.
187. Khallaf, A.N. Using AI to Help Reduce the Effect of Global Warming. *Power Syst. Technol.* **2024**, *48*, 1927–1947. [CrossRef]
188. Wang, Y.; Ho, I.W.-H.; Chen, Y.; Wang, Y.; Lin, Y. Real-Time Water Quality Monitoring and Estimation in AIoT for Freshwater Biodiversity Conservation. *IEEE Internet Things J.* **2021**, *9*, 14366–14374. [CrossRef]
189. Zulkifli, C.Z.; Garfan, S.; Talal, M.; Alamoodi, A.H.; Alamleh, A.; Ahmaro, I.Y.; Sulaiman, S.; Ibrahim, A.B.; Zaidan, B.B.; Ismail, A.R. IoT-Based Water Monitoring Systems: A Systematic Review. *Water* **2022**, *14*, 3621. [CrossRef]
190. Mustafa, H.M.; Mustapha, A.; Hayder, G.; Salisu, A. Applications of IoT and Artificial Intelligence in Water Quality Monitoring and Prediction: A Review. In Proceedings of the 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 20–22 January 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 968–975. [CrossRef]
191. Bradley, R.W.; Sprague, J.B. The Influence of pH, Water Hardness, and Alkalinity on the Acute Lethality of Zinc to Rainbow Trout (*Salmo gairdneri*). *Can. J. Fish. Aquat. Sci.* **1985**, *42*, 731–736. [CrossRef]
192. Sigdel, B. *Water Quality Measuring Station: pH, Turbidity and Temperature Measurement*; Technical Report; Metropolia University of Applied Sciences: Helsinki, Finland, 2017.
193. Lloyd, D.S.; Koenings, J.P.; Laperriere, J.D. Effects of Turbidity in Fresh Waters of Alaska. *N. Am. J. Fish. Manag.* **1987**, *7*, 18–33. [CrossRef]
194. Gorde, S.P.; Jadhav, M.V. Assessment of Water Quality Parameters: A Review. *J. Eng. Res. Appl.* **2013**, *3*, 2029–2035.
195. Kannel, P.R.; Lee, S.; Lee, Y.-S.; Kanel, S.R.; Khan, S.P. Application of Water Quality Indices and Dissolved Oxygen as Indicators for River Water Classification and Urban Impact Assessment. *Environ. Monit. Assess.* **2007**, *132*, 93–110. [CrossRef] [PubMed]
196. Best, M.A.; Wither, A.W.; Coates, S. Dissolved Oxygen as a Physico-Chemical Supporting Element in the Water Framework Directive. *Mar. Pollut. Bull.* **2007**, *55*, 53–64. [CrossRef]
197. Nazeer, S.; Hashmi, M.Z.; Malik, R.N. Heavy Metals Distribution, Risk Assessment and Water Quality Characterization by Water Quality Index of the River Soan, Pakistan. *Ecol. Indic.* **2014**, *43*, 262–270. [CrossRef]
198. Mokarram, M.; Saber, A.; Sheykhi, V. Effects of Heavy Metal Contamination on River Water Quality Due to Release of Industrial Effluents. *J. Clean. Prod.* **2020**, *277*, 123380. [CrossRef]
199. Rabeh, S.A. Bacteria and Viruses in the Nile. In *The Nile*; Dumont, H.J., Ed.; Monographiae Biologicae; Springer: Dordrecht, The Netherlands, 2009; Volume 89, pp. 407–429, ISBN 978-1-4020-9725-6.
200. Lin, J.; Ganesh, A. Water Quality Indicators: Bacteria, Coliphages, Enteric Viruses. *Int. J. Environ. Health Res.* **2013**, *23*, 484–506. [CrossRef]
201. Working EIFAC. Water Quality Criteria for European Freshwater Fish—Extreme pH Values and Inland Fisheries. *Water Res.* **1969**, *3*, 593–611. [CrossRef]
202. Banna, M.H.; Najjaran, H.; Sadiq, R.; Imran, S.A.; Rodriguez, M.J.; Hoorfar, M. Miniaturized Water Quality Monitoring pH and Conductivity Sensors. *Sens. Actuators Chem.* **2014**, *193*, 434–441. [CrossRef]
203. Bilotta, G.S.; Brazier, R.E. Understanding the Influence of Suspended Solids on Water Quality and Aquatic Biota. *Water Res.* **2008**, *42*, 2849–2861. [CrossRef]
204. Yalaletdinova, A.V.; Beloliptsev, I.I.; Galimova, Y.O.; Vozhdaeva, M.Y.; Kantor, E.A. Probability Analysis of Water Quality by Turbidity. *Iop Conf. Ser. Earth Environ. Sci.* **2019**, *315*, 062019. [CrossRef]
205. Gunda, N.S.; Gautam, S.; Mitra, S. Artificial Intelligence for Water Quality Monitoring. In *ECS Meeting Abstracts*; The Electrochemical Society: Pennington, NJ, USA, 2018; p. 1997.
206. Biraghi, C.A.; Loftian, M.; Carrion, D.; Brovelli, M.A. AI in Support to Water Quality Monitoring. In Proceedings of the XXIV ISPRS Congress, Virtual, 5–9 July 2021.
207. Strobl, R.O.; Robillard, P.D. Artificial Intelligence Technologies in Surface Water Quality Monitoring. *Water Int.* **2006**, *31*, 198–209. [CrossRef]

208. Tung, T.M.; Yaseen, Z.M. A Survey on River Water Quality Modelling Using Artificial Intelligence Models: 2000–2020. *J. Hydrol.* **2020**, *585*, 124670. [CrossRef]
209. Nvs, B.; Saranya, P.L. Water Pollutants Monitoring Based on Internet of Things. In *Inorganic Pollutants in Water*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 371–397. [CrossRef]
210. Singh, Y.; Walingo, T. Smart Water Quality Monitoring with IoT Wireless Sensor Networks. *Sensors* **2024**, *24*, 2871. [CrossRef] [PubMed]
211. Hossain, M.S.; Rahman, M.K.; Dalim, H.M. Leveraging AI for Real-Time Monitoring and Prediction of Environmental Health Hazards: Protecting Public Health in the USA. *Rev. Intell. Artif. Med. Appl.* **2024**, *15*, 1117–1145.
212. Mohanty, A.; Mohanty, S.K.; Mohapatra, A.G. Real-Time Monitoring and Fault Detection in AI-Enhanced Wastewater Treatment Systems. In *The AI Cleanse: Transforming Wastewater Treatment Through Artificial Intelligence*; Garg, M.C., Ed.; Springer Water; Springer: Cham, Switzerland, 2023; pp. 165–199, ISBN 978-3-031-67236-1.
213. Singh, S.; Kumar, A.; Prasad, A.; Bharadwaj, N. IoT Based Water Quality Monitoring System. In Proceedings of the IRFIC 2016, New Delhi, India, 10 January 2016; Institute of Research and Journals (IRAJ): Bhubaneswar, India, 2016.
214. Sugiharto, W.H.; Susanto, H.; Prasetijo, A.B. Real-Time Water Quality Assessment via IoT: Monitoring pH, TDS, Temperature, and Turbidity. *J. Inf. Intell. Eng. Appl.* **2023**, *28*, 823–831. [CrossRef]
215. Yunfeng, L.; Tianpei, Z. A Design of Dissolved Oxygen Monitoring System Based on NB-IoT. In Proceedings of the 2019 International Conference on Smart Grid and Electrical Automation (ICSGEA), Zhangjiajie, China, 7–9 June 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 98–101. [CrossRef]
216. Wang, L.Y.K.; Bong, J.W.; Eng, Y.X.; Wong, W.S. IoT Based Wastewater Dissolved Oxygen and Total Dissolved Solids Monitoring with Data Analytics. In Proceedings of the 2024 11th International Conference on Future Internet of Things and Cloud (FiCloud), Istanbul, Turkey, 26–28 August 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 98–103.
217. Bria, A.R.; Cerro, G.; Ferdinandi, M.; Marrocco, C.; Molinara, M. An IoT-Ready Solution for Automated Recognition of Water Contaminants. *Pattern Recognit. Lett.* **2020**, *135*, 188–195. [CrossRef]
218. Pamula, A.S.; Ravilla, A.; Madiraju, S.V.H. Applications of the Internet of Things (IoT) in Real-Time Monitoring of Contaminants in the Air, Water, and Soil. *Eng. Proc.* **2022**, *27*, 26. [CrossRef]
219. Ooko, S.O.; Rweyemamu, E. Monitoring and Predicting African Rural Household Air Pollution Using Internet of Things and Artificial Intelligence. *Pan-Afr. J. Health Environ. Sci.* **2024**, *3*, 59. [CrossRef]
220. Konar, M. Integrating IoT and AI for Automated Crop with Smart Urban Farming. In Proceedings of the 2024 International Conference on Emerging Innovations and Advanced Computing (INNOCOMP), Sonipat, India, 25–26 May 2024. [CrossRef]
221. 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] [PubMed]
222. Alselek, M.; Alcaraz-Calero, J.; Wang, Q. Dynamic AI-IoT: Enabling Updatable AI Models in Ultralow-Power 5G IoT Devices. *IEEE Internet Things J.* **2024**, 14192. [CrossRef]
223. Rathoure, A.K.; Rathaur, A.K. AI and IoT in Biodiversity Assessment. *J. Aquac. Mar. Biol.* **2024**, *13*, 108. [CrossRef]
224. Gupta, S.; Hasan, W.; Singh, S.; Kumar, D.; Ansari, M.J.; Nisar, S. (Eds.) Ethical Considerations of AI and IoT in Farming, Chapter 16. In *Book Agriculture 4.0 Smart Farming with IoT and Artificial Intelligence*; CRC Press: London, UK, 2024. [CrossRef]
225. Yiou, P.; Baert, E.; Loutre, M.F. Spectral Analysis of Climate Data. *Surv. Geophys.* **1996**, *17*, 619–663. [CrossRef]
226. Mitchell, T.D.; Hulme, M.; New, M. Climate Data for Political Areas. *Area* **2002**, *34*, 109–112. [CrossRef]
227. Overpeck, J.T.; Meehl, G.A.; Bony, S.; Easterling, D.R. Climate Data Challenges in the 21st Century. *Science* **2011**, *331*, 700–702. [CrossRef] [PubMed]
228. Box, J.E.; Colgan, W.T.; Christensen, T.R.; Schmidt, N.M.; Lund, M.; Parmentier, F.-J.W.; Brown, R.; Bhatt, U.S.; Euskirchen, E.S.; Romanovsky, V.E. Key Indicators of Arctic Climate Change: 1971–2017. *Environ. Res. Lett.* **2019**, *14*, 045010. [CrossRef]
229. Kenney, M.A.; Janetos, A.C.; Gerst, M.D. A Framework for National Climate Indicators. *Clim. Chang.* **2020**, *163*, 1705–1718. [CrossRef]
230. Kjellström, E.; Boberg, F.; Castro, M.; Christensen, J.H.; Nikulin, G.; Sánchez, E. Daily and Monthly Temperature and Precipitation Statistics as Performance Indicators for Regional Climate Models. *Clim. Res.* **2010**, *44*, 135–150. [CrossRef]
231. Bowler, D.; Böhning-Gaese, K. Improving the Community-Temperature Index as a Climate Change Indicator. *PLoS ONE* **2017**, *12*, e0184275. [CrossRef] [PubMed]
232. Wypych, A. Twentieth Century Variability of Surface Humidity as the Climate Change Indicator in Kraków (Southern Poland). *Theor. Appl. Climatol.* **2010**, *101*, 475–482. [CrossRef]
233. Matthews, T. Humid Heat and Climate Change. *Prog. Phys. Geogr. Earth Environ.* **2018**, *42*, 391–405. [CrossRef]

234. Groisman, P.Y.; Karl, T.R.; Easterling, D.R.; Knight, R.W.; Jamason, P.F.; Hennessy, K.J.; Suppiah, R.; Page, C.M.; Wibig, J.; Fortuniak, K.; et al. Changes in the Probability of Heavy Precipitation: Important Indicators of Climatic Change. In *Weather and Climate Extremes*; Karl, T.R., Nicholls, N., Ghazi, A., Eds.; Springer: Dordrecht, The Netherlands, 1999; pp. 243–283, ISBN 978-90-481-5223-0.
235. Pal, I.; Al-Tabbaa, A. Trends in Seasonal Precipitation Extremes–An Indicator of 'Climate Change' in Kerala, India. *J. Hydrol.* **2009**, *367*, 62–69. [CrossRef]
236. Alpert, P.; Mandel, M. Wind Variability—An Indicator for a Mesoclimatic Change in Israel. *J. Clim. Appl. Meteorol.* **1986**, *25*, 1568–1576. [CrossRef]
237. Dewitte, S.; Cornelis, J.P.; Müller, R.; Munteanu, A. Artificial Intelligence Revolutionises Weather Forecast, Climate Monitoring and Decadal Prediction. *Remote Sens.* **2021**, *13*, 3209. [CrossRef]
238. Rahman, A.; Saha, R.; Goswami, D.; Mintoo, A.A. Climate Data Management Systems: Systematic Review of Analytical Tools for Informing Policy Decisions. *Front. Appl. Eng. Technol.* **2024**, *1*, 1–21. [CrossRef]
239. Salam, A. Internet of Things for Environmental Sustainability and Climate Change. In *Internet of Things for Sustainable Community Development*; Internet of Things; Springer International Publishing: Cham, Switzerland, 2024; pp. 33–69, ISBN 978-3-031-62161-1.
240. Gutman, S.I.; Benjamin, S.G. The Role of Ground-Based GPS Meteorological Observations in Numerical Weather Prediction. *GPS Solut.* **2001**, *4*, 16–24. [CrossRef]
241. De Sanctis, M.; Cianca, E.; Araniti, G.; Bisio, I.; Prasad, R. Satellite Communications Supporting Internet of Remote Things. *IEEE Internet Things J.* **2015**, *3*, 113–123. [CrossRef]
242. Chen, Y.; Zhang, M.; Li, X.; Che, T.; Jin, R.; Guo, J.; Yang, W.; An, B.; Nie, X. Satellite-Enabled Internet of Remote Things Network Transmits Field Data from the Most Remote Areas of the Tibetan Plateau. *Sensors* **2022**, *22*, 3713. [CrossRef] [PubMed]
243. Almalki, F.A.; Soufiene, B.O.; Alsamhi, S.H.; Sakli, H. A Low-Cost Platform for Environmental Smart Farming Monitoring System Based on IoT and UAVs. *Sustainability* **2021**, *13*, 5908. [CrossRef]
244. Revathi, S.; Ansari, A.; Susmi, S.J.; Madhavi, M.; Gunavathie, M.A.; Sudhakar, M. Integrating Machine Learning–IoT Technologies Integration for Building Sustainable Digital Ecosystems. In *Advances in Computational Intelligence and Robotics*; Kajla, T., Kansra, P., Singh, N., Eds.; IGI Global: Hershey, PA, USA, 2024; pp. 259–291, ISBN 979-8-3693-2432-5.
245. Ali, M.A.; Dhanaraj, R.K.; Nayyar, A. A High Performance-Oriented AI-Enabled IoT-Based Pest Detection System Using Sound Analytics in Large Agricultural Field. *Microprocess. Microsyst.* **2023**, *103*, 104946. [CrossRef]
246. Farhaoui, Y.; El Allaoui, A. Sustainability in the Internet of Things: Insights, Scope, and AI-Driven Optimized Water Management with Big Data Integration. In *Artificial Intelligence, Big Data, IOT and Block Chain in Healthcare: From Concepts to Applications*; Farhaoui, Y., Ed.; Information Systems Engineering and Management; Springer: Cham, Switzerland, 2024; Volume 6, pp. 468–475, ISBN 978-3-031-65017-8.
247. Maraveas, C.; Piromalis, D.; Arvanitis, K.G.; Bartzanas, T.; Loukatos, D. Applications of IoT for Optimized Greenhouse Environment and Resources Management. *Comput. Electron. Agric.* **2022**, *198*, 106993. [CrossRef]
248. Bale, A.S.; William, P.; Kondekar, V.H.; Sanamdikar, S.; Joshi, P.; Nigam, P.; Savadatti, M.B. Harnessing AI and IoT for Optimized Renewable Energy Integration and Resource Conservation. *Libr. Prog. Int.* **2024**, *44*, 1412–1426.
249. Bianchi, O.; Putro, H.P. Artificial Intelligence in Environmental Monitoring: Predicting and Managing Climate Change Impacts. *Int. Trans. Artif. Intell. (ITALIC)* **2024**, *3*, 85–96. [CrossRef]
250. Park, J.; Kim, K.T.; Lee, W.H. Recent Advances in Information and Communications Technology (ICT) and Sensor Technology for Monitoring Water Quality. *Water* **2020**, *12*, 510. [CrossRef]
251. Kaginalkar, A.; Kumar, S.; Gargava, P.; Niyogi, D. Review of Urban Computing in Air Quality Management as Smart City Service: An Integrated IoT, AI, and Cloud Technology Perspective. *Urban Clim.* **2021**, *39*, 100972. [CrossRef]
252. Chui, K.T.; Lytras, M.D.; Visvizi, A. Energy Sustainability in Smart Cities: Artificial Intelligence, Smart Monitoring, and Optimization of Energy Consumption. *Energies* **2018**, *11*, 2869. [CrossRef]
253. Kumar, D.; Shekhar, S.; Tewary, T. Data Analytics and Artificial Intelligence in Earth Resource Management. In *Data Analytics and Artificial Intelligence for Earth Resource Management*; Elsevier: Amsterdam, The Netherlands, 2025; pp. 1–17.
254. Rane, N.; Choudhary, S.; Rane, J. Leading-Edge Artificial Intelligence (AI), Machine Learning (ML), Blockchain, and Internet of Things (IoT) Technologies for Enhanced Wastewater Treatment Systems. *SSRN Electron. J.* **2023**. [CrossRef]
255. Amirian, H.; Dalvand, K.; Ghiasvand, A. Seamless Integration of Internet of Things, Miniaturization, and Environmental Chemical Surveillance. *Environ. Monit. Assess.* **2024**, *196*, 582. [CrossRef] [PubMed]
256. Salman, O.; Elhajj, I.; Kayssi, A.; Chehab, A. Edge Computing Enabling the Internet of Things. In Proceedings of the 2015 IEEE 2nd World Forum on Internet of Things (WF-IoT), Milan, Italy, 14–16 December 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 603–608.
257. Hassan, N.; Gillani, S.; Ahmed, E.; Yaqoob, I.; Imran, M. The Role of Edge Computing in Internet of Things. *IEEE Commun. Mag.* **2018**, *56*, 110–115. [CrossRef]

258. Kaur, S.; Kumar, R.; Singh, K.; Huang, Y.L. Leveraging Artificial Intelligence for Enhanced Sustainable Energy Management. *J. Sustain. Energy* **2024**, *3*, 1–20. [CrossRef]
259. Whig, P.; Remala, R.; Mudunuru, K.R.; Quraishi, S.J. Integrating AI and Quantum Technologies for Sustainable Supply Chain Management. In *Quantum Computing and Supply Chain Management: A New Era of Optimization*; IGI Global: Hershey, PA, USA, 2024; pp. 267–283.
260. Bukhari, S.A.S.; Shafi, I.; Ahmad, J.; Butt, H.T.; Khurshaid, T.; Ashraf, I. Enhancing Flood Monitoring and Prevention Using Machine Learning and IoT Integration. *Nat. Hazards* **2024**. [CrossRef]
261. Indrakumari, R.; Sriramulu, S.; Partheeban, N.; Rajavel, R. FloodWatch: Suggesting an IoT-Driven Flood Monitoring and Early Warning System for the Flood-Prone Cuddalore District in the Indian State of Tamilnadu. In *Digital Twin Technology and Applications*; Auerbach Publications: Boca Raton, FL, USA, 2024; pp. 374–387.
262. Ukoba, K.; Olatunji, K.O.; Adeoye, E.; Jen, T.-C.; Madyira, D.M. Optimizing Renewable Energy Systems through Artificial Intelligence: A Review and Future Prospects. *Energy Environ.* **2024**, *35*, 3833–3879. [CrossRef]
263. Wen, X.; Shen, Q.; Zheng, W.; Zhang, H. AI-Driven Solar Energy Generation and Smart Grid Integration: A Holistic Approach to Enhancing Renewable Energy Efficiency. *Int. J. Innov. Res. Eng. Manag.* **2024**, *11*, 55–66. [CrossRef]
264. Teh, H.Y.; Kempa-Liehr, A.W.; Wang, K.I.-K. Sensor Data Quality: A Systematic Review. *J. Big Data* **2020**, *7*, 11. [CrossRef]
265. Concas, F.; Mineraud, J.; Lagerspetz, E.; Varjonen, S.; Liu, X.; Puolamäki, K.; Nurmi, P.; Tarkoma, S. Low-Cost Outdoor Air Quality Monitoring and Sensor Calibration: A Survey and Critical Analysis. *ACM Trans. Sens. Netw.* **2021**, *17*, 1–44. [CrossRef]
266. Deekshith, A. Data Engineering for AI: Optimizing Data Quality and Accessibility for Machine Learning Models. *Int. J. Manag. Educ. Sustain. Dev.* **2021**, *4*, 1–33.
267. Gade, K.R. Data Quality Metrics for the Modern Enterprise: A Data Analytics Perspective. *MZ J. Artif. Intell.* **2024**, *1*.
268. Qazi, S.; Khawaja, B.A.; Farooq, Q.U. IoT-Equipped and AI-Enabled Next Generation Smart Agriculture: A Critical Review, Current Challenges and Future Trends. *IEEE Access* **2022**, *10*, 21219–21235. [CrossRef]
269. Wang, X.; Li, X.; Leung, V.C.M. Artificial Intelligence-Based Techniques for Emerging Heterogeneous Network: State of the Arts, Opportunities, and Challenges. *IEEE Access* **2015**, *3*, 1379–1391. [CrossRef]
270. Pankaj, D.A.B. The Internet of Things and Cyber-Physical Systems: AI-Enhanced Interoperability and Efficiency. *Int. IT J. Res.* **2024**, *2*, 53–63.
271. Dave, D.M.K.; Mittapally, B.K. Data Integration and Interoperability in IoT: Challenges, Strategies and Future Direction. *Int. J. Comput. Eng. Technol. (IJCET)* **2024**, *15*, 45–60.
272. Tawalbeh, L.; Muheidat, F.; Tawalbeh, M.; Quwaider, M. IoT Privacy and Security: Challenges and Solutions. *Appl. Sci.* **2020**, *10*, 4102. [CrossRef]
273. Marengo, A. The Future of AI in IoT: Emerging Trends in Intelligent Data Analysis and Privacy Protection. 2023. Available online: https://www.researchgate.net/publication/377087354_The_Future_of_AI_in_IoT_Emerging_Trends_in_Intelligent_Data_Analysis_and_Privacy_Protection?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6Il9kaXJlY3QiLCJwYWdlIjoicHVibGljYXRpb24ifX0 (accessed on 27 December 2024).
274. Karale, A. The Challenges of IoT Addressing Security, Ethics, Privacy, and Laws. *Internet Things* **2021**, *15*, 100420. [CrossRef]
275. Čolaković, A.; Hadžialić, M. Internet of Things (IoT): A Review of Enabling Technologies, Challenges, and Open Research Issues. *Comput. Netw.* **2018**, *144*, 17–39. [CrossRef]
276. Falola, P.B.; Adeniyi, A.E.; Madamidola, O.A.; Awotunde, J.B.; Olukiran, O.A.; Akinola, S.O. Artificial Intelligence in Agriculture: The Potential for Efficiency and Sustainability, with Ethical Considerations. In *Exploring Ethical Dimensions of Environmental Sustainability and Use of AI*; IGI Global Scientific Publishing: Hershey, PA, USA, 2024; pp. 307–329, ISBN 979-8-3693-0892-9.
277. Akter, S. Ethical AI Development for Sustainable Enterprises: A Review of Integrating Responsible AI with IoT and Enterprise Systems. *J. Artif. Intell. Gen. Sci. (JAIGS)* **2024**, *6*, 94–125. [CrossRef]
278. Perera, A.T.D.; Kamalaruban, P. Applications of Reinforcement Learning in Energy Systems. *Renew. Sustain. Energy Rev.* **2021**, *137*, 110618. [CrossRef]
279. Dang, T.; Liu, J. Adaptive Water Environment Optimization Strategy Based on Reinforcement Learning. *Comput.-Aided Des. Appl.* **2024**, *21*, 1–18. [CrossRef]
280. Fan, Z.; Yan, Z.; Wen, S. Deep Learning and Artificial Intelligence in Sustainability: A Review of SDGs, Renewable Energy, and Environmental Health. *Sustainability* **2023**, *15*, 13493. [CrossRef]
281. Khanmohammadi, A.; Jalili Ghazizadeh, A.; Hashemi, P.; Afkhami, A.; Arduini, F.; Bagheri, H. An Overview to Electrochemical Biosensors and Sensors for the Detection of Environmental Contaminants. *J. Iran. Chem. Soc.* **2020**, *17*, 2429–2447. [CrossRef]
282. Manzano, L.G.; Boukabache, H.; Danzeca, S.; Heracleous, N.; Murtas, F.; Perrin, D.; Pirc, V.; Alfaro, A.R.; Zimmaro, A.; Silari, M. An IoT LoRaWAN Network for Environmental Radiation Monitoring. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–12. [CrossRef]
283. Wang, L.; Yang, H. Design of Smart City Environment Monitoring and Optimisation System Based on NB-IoT Technology. *Int. J. Inf. Commun. Technol.* **2024**, *25*, 47–61. [CrossRef]

284. Niu, L. Design of Intelligent Agricultural Environmental Big Data Collection System Based on ZigBee and NB-IoT. In Proceedings of the 2023 IEEE 2nd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA), Beijing, China, 25–27 August 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1299–1304.
285. Jabbar, W.A.; Subramaniam, T.; Ong, A.E.; Shu'Ib, M.I.; Wu, W.; De Oliveira, M.A. LoRaWAN-Based IoT System Implementation for Long-Range Outdoor Air Quality Monitoring. *Internet Things* **2022**, *19*, 100540. [CrossRef]
286. Phupattanasilp, P.; Tong, S.-R. Augmented Reality in the Integrative Internet of Things (AR-IoT): Application for Precision Farming. *Sustainability* **2019**, *11*, 2658. [CrossRef]
287. Filipova, M.; Gospodinov, G.; Gotsev, L.; Kovatcheva, E.; Jekov, B. Quantum Computing Applications for Addressing Global Warming and Pollution: A Comprehensive Analysis. In Proceedings of the 15th International Scientific and Practical Conference "Environment. Technology. Resources", Rezekne, Latvia, 19–20 June 2024; Volume 1, pp. 154–160.
288. Vanderbilt, K.L.; Lin, C.-C.; Lu, S.-S.; Kassim, A.R.; He, H.; Guo, X.; Gil, I.S.; Blankman, D.; Porter, J.H. Fostering Ecological Data Sharing: Collaborations in the International Long Term Ecological Research Network. *Ecosphere* **2015**, *6*, 1–18. [CrossRef]
289. Fascista, A. Toward Integrated Large-Scale Environmental Monitoring Using WSN/UAV/Crowdsensing: A Review of Applications, Signal Processing, and Future Perspectives. *Sensors* **2022**, *22*, 1824. [CrossRef] [PubMed]
290. Okafor, N. Advances and Challenges in IoT Sensors Data Handling and Processing in Environmental Monitoring Systems. 2023. Available online: https://www.techrxiv.org/users/679289/articles/682738-advances-and-challenges-in-iot-sensors-data-handling-and-processing-in-environmental-monitoring-systems (accessed on 27 December 2024 ).
291. Gouveia, C.; Fonseca, A.; Câmara, A.; Ferreira, F. Promoting the Use of Environmental Data Collected by Concerned Citizens through Information and Communication Technologies. *J. Environ. Manag.* **2004**, *71*, 135–154. [CrossRef] [PubMed]
292. Conrad, C.C.; Hilchey, K.G. A Review of Citizen Science and Community-Based Environmental Monitoring: Issues and Opportunities. *Environ. Monit. Assess.* **2011**, *176*, 273–291. [CrossRef]
293. Berigüete, F.E.; Santos, J.S.; Rodriguez Cantalapiedra, I. Digital Revolution: Emerging Technologies for Enhancing Citizen Engagement in Urban and Environmental Management. *Land* **2024**, *13*, 1921. [CrossRef]
294. Cao, F.; Jian, Y. The Role of Integrating AI and VR in Fostering Environmental Awareness and Enhancing Activism among College Students. *Sci. Total Environ.* **2024**, *908*, 168200. [CrossRef]
295. Asuquo, A.; Agbor, C.; Bassey, A.; Omoogun, M. Artificial Intelligence (AI) Innovations for Sustainable Educational Institutions: Enhancing Efficiency and Environmental Responsibility. *Int. J. Innov. Technol. Integr. Educ.* **2024**, *7*, 84–101.
296. Ho, K.T.M.; Chen, K.-C.; Lee, L.; Burt, F.; Yu, S.; Lee, P.-H. Quantum Computing for Climate Resilience and Sustainability Challenges. 2024. Available online: https://arxiv.org/abs/2407.16296 (accessed on 28 December 2024).
297. Arabelli, R.; Boddepalli, E.; Buradkar, M.; Goriparti, N.V.S.; Chakravarthi, M.K. IoT-Enabled Environmental Monitoring System Using AI. In Proceedings of the 2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 9–10 May 2024; pp. 1–6. [CrossRef]
298. Arowolo, M.E.; Aaron, W.C.; Eteng, U.S.; Divine, I.L.O.H.; Aguma, C.P.; Olagunju, A.O. Integrating AI-Enhanced Remote Sensing Technologies with IoT Networks for Precision Environmental Monitoring and Predictive Ecosystem Management. *World J. Adv. Res. Rev.* **2024**, *23*, 2156–2166. [CrossRef]
299. Manongga, D.; Rahardja, U.; Sembiring, I.; Aini, Q.; Wahab, A. Improving the Air Quality Monitoring Framework Using Artificial Intelligence for Environmentally Conscious Development. *HighTech Innov. J.* **2024**, *5*, 794–813. [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

*Electronics* Editorial Office
E-mail: electronics@mdpi.com
www.mdpi.com/journal/electronics




Disclaimer/Publisher's Note: The title and front matter of this reprint are at the discretion of the Guest 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 978-3-7258-9374-4