

Article

An Innovative Index for Evaluating Urban Vulnerability on Pandemic Using LambdaMART Algorithm

Yuming Lin and Zhenjiang Shen * 

Division of Environmental Design, Kanazawa University, Kanazawa 920-1192, Japan;
linyuming@stu.kanazawa-u.ac.jp

* Correspondence: shenzhe@se.kanazawa-u.ac.jp

Abstract: The COVID-19 pandemic has significantly changed urban life and increased attention has been paid to the pandemic in discussions of urban vulnerability. There is a lack of methods to incorporate dynamic indicators such as urban vitality into evaluations of urban pandemic vulnerability. In this research, we use machine learning to establish an urban Pandemic Vulnerability Index (PVI) that measures the city's vulnerability to the pandemic and takes dynamic indicators as an important aspect of this. The proposed PVI is constructed using 140 statistic variables and 10 dynamic variables, using data from 47 prefectures of Japan. Factor Analysis is used to extract factors from variables that may affect city vulnerability, and the LambdaMART algorithm is used to aggregate factors and predict vulnerability. The results show that the proposed PVI can predict the relative seriousness of the COVID-19 pandemic in two weeks with a precision of more than 0.71, which is meaningful for taking controlling measures in advance and shaping the society's response. Further analysis revealed the key factors affecting urban pandemic vulnerability, including city size, transit station vitality, and medical facilities, emphasizing precautions for public transport systems and new planning concepts such as the compact city. This research explores the application of machine learning techniques in the indicator establishment and incorporates dynamic factors into vulnerability assessments, which contribute to improvements in urban vulnerability assessments and the planning of sustainable cities while facing the challenges of the COVID-19 pandemic.

Keywords: urban vulnerability; pandemic vulnerability index; COVID-19; factor analysis; LambdaMART



Citation: Lin, Y.; Shen, Z. An Innovative Index for Evaluating Urban Vulnerability on Pandemic Using LambdaMART Algorithm. *Sustainability* **2022**, *14*, 5053. <https://doi.org/10.3390/su14095053>

Academic Editors: Sina Shahab, Marica Cassarino and Sara Biscaya

Received: 20 March 2022

Accepted: 20 April 2022

Published: 22 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Against the background of the COVID-19 pandemic, the concept of resistance and vulnerability has drawn more and more attention [1]. Urban vulnerability represents a state of being likely to be influenced by natural disasters, including earthquakes, typhoons, and floods. When natural disasters strike, high-vulnerability cities and high-vulnerability areas in the city will be more likely to be harmed, inflicting suffering on vulnerable groups and exacerbating existing inequalities [2]. Recent studies show that pandemics such as COVID-19 should also be regarded as natural disasters and be included in the discussion of urban vulnerability [3,4].

In urban planning, many urban factors might affect urban vulnerability, such as excessive population density, low-quality housing, inadequate infrastructure, and environmental degradation [2,5]. The differences between urban factors bring different levels of urban vulnerability, leading to different performances in response to natural disasters. For example, income inequality will affect a region's vulnerability to flooding [6]. The COVID-19 pandemic also shows regional differences in its spread, bringing more damage to vulnerable countries and regions [7,8]. UN-Habitat pointed out that the Asia-Pacific region could be the most susceptible due to its fast urbanization rate, with one-third of urban dwellers in slums or slum-like conditions [9]. Given the still-raging pandemic, assessing urban vulnerability to the pandemic is an urgent task.

Urban vulnerability is usually defined as the degree to which a system is susceptible to, and unable to cope with, hazards or stresses [2]. As an extension of urban vulnerability, we define urban pandemic vulnerability as the extent to which an urban system is susceptible to sustaining damage from the pandemic. Therefore, improving urban pandemic vulnerability implies reducing exposure to pandemics and increasing the capacity to resist pandemic damage. Assessing urban pandemic vulnerability will allow for planners and policymakers to inspect the potential impact of the urban built environment on health and expose inequalities between urban areas [10,11].

Although some related studies have assessed urban pandemic vulnerability [3,12,13], there have not been suitable methods considering dynamic factors, such as urban vitality and prevention measures. Additionally, the limited advances in urban dynamics in urban vulnerability assessments are considered an important problem for urban strategic planning [10]. Therefore, this research would like to build a Pandemic Vulnerability Index (PVI) to evaluate an urban area's vulnerability to the pandemic and consider the impacts of dynamic factors, which can improve the capacity to capture the dynamic nature of the urban vulnerability, advance research related to urban vulnerability dynamics, and benefit assessment-based urban planning in the post-COVID-19 era.

2. Literature Review

Unlike natural disasters such as earthquakes and typhoons, which can suddenly wreak havoc on urban facilities, pandemics such as COVID-19 threaten cities by harming residents' health, stopping facilities from functioning, and disrupting daily urban life [14]. On the one hand, the pressure brought on by the pandemic will be transmitted to all aspects of urban life through various paths [15], making it more challenging to identify the possible vulnerable link. On the other hand, the pandemic gradually causes damage as it spreads, which is dynamic and occurs over a relatively long period [16]. This means that the risk of exposure and damage also dynamically changes along with the pandemic spread and the city's reaction.

Since the COVID-19 outbreak in 2020, there has been some research on pandemic vulnerabilities. Mishra, Gayen and Haque [3] examined four major cities of India, devised a COVID Vulnerability Index with carefully selected indicators, and analyzed why social distancing and lockdowns failed in vulnerable slums. Prieto, Malagón, Gomez and León [12] proposed an Urban Vulnerability Assessment methodology to investigate the various vulnerability factors related to pandemics and aggregate them into a vulnerability index using the data from Bogotá, Colombia. Shi, Liao, Li and Su [13] employed the crisp-set qualitative comparative analysis method to explore possible causal condition combination paths that affect community resilience to the pandemic in Wuhan, China, showing three condition configurations that were vulnerable to the pandemic, including communities populated by disadvantaged populations. These pieces of research provide an essential analytical framework for identifying possible vulnerability factors in cities, and validate the feasibility of extracting vulnerability factors from qualitative or quantitative data using methods such as Factor Analysis.

However, the dynamic pressure of the pandemic has created new difficulties for researchers. Taking Japan as an example, in the third wave of the pandemic in January 2021, the most densely populated Tokyo metropolitan area had the highest number of new infections [17]. However, in the fourth wave in June 2021, Okinawa, which performed relatively better in the last wave, had the highest number of infections and was short of medical resources [18]. This situation indicates that it is not enough to only focus on the inherent factors of cities, which led to the examination of dynamic urban indicators [16]. Practical methods that can investigate dynamic variables such as urban vitality and prevention measures such as social distancing are needed when conducting an urban vulnerability assessment. Machine learning techniques, which have gained popularity in recent years, provide some proven approaches: Zawbaa, et al. [19] used the Multi-Layer Perceptron to model the spread of COVID-19 and verified the impact of social distancing; and

Pan, et al. [20] used Random Forest to capture pandemic dynamics and make time-series predictions, and further offered optimal solutions to minimize the growth of confirmed cases and deaths through NSGA-II. These pieces of research illustrate that machine learning technology can capture the dynamic spread of the pandemic and use empirical data to verify the results.

Therefore, this research aims to establish a composite index to evaluate urban vulnerability to the pandemic. The proposed composite index is innovative in considering the impact of dynamic factors on urban vulnerability, which can make up for the insufficient advances in the urban dynamics research line of urban vulnerability assessments [10]. The Factor Analysis (FA) will extract essential factors from the available indicators, and a machine learning LambdaMART algorithm is used to combine these factors into a Pandemic Vulnerability Index (PVI). PVI's prediction ability will be verified on empirical data, and the critical characteristics that influence urban pandemic vulnerability will be examined through feature importance and dependence analysis. The proposed PVI is expected to dynamically identify vulnerable regions and remind decision-makers in the corresponding region to take preventive measures such as social distancing or expanding healthcare capacity in advance. The corresponding analysis is expected to reveal the key influencing factors, such as urban sprawl, that should be carefully considered in future urban planning. We believe that a pandemic vulnerability index that includes dynamic factors can refine the framework for urban vulnerability assessments and contribute to flexible and accurate city planning and policies in the post-COVID-19 era. Machine learning techniques, including FA or LambdaMART, are a promising method.

3. Materials and Methods

3.1. Situation of Japan

As a democratic government that emphasizes local autonomy, Japan's prefectures show significant differences among regions. From the urban foundation perspective, the cities of the three major metropolitan areas around Tokyo, Osaka, and Nagoya have formed a distinctive urban form with high population density and a relatively complete infrastructure, differing from other prefectures. Different prefectures have made different policies in response to the pandemic due to their different pandemic situations and economic considerations. For example, Hokkaido's local government declared measures for the pandemic as early as 28 February 2020, while the national state of emergency was declared on 7 April 2020. The differences in urban infrastructure, residents' actions, and government policy will affect the cities' ability to counter the pandemic and become the basis for establishing and verifying the cities' vulnerability index.

3.2. Data Source and Software

The Japanese government has released various databases to fight the COVID-19 pandemic and support related research. In addition, many Internet service providers, such as Google or NTT, have also released a series of data related to human activities during the pandemic. The data used in this research are from the public database, including:

- Digital National Land Information; [21]
- COVID-19 Trends and Current Situation; [22]
- Status of Prefectural Medical Care Provision System; [23]
- Community Mobility Reports; [24]
- Coronavirus Support Site. [25]

These databases cover necessary information for each prefecture, such as population, GDP, medical facilities, and the dynamic changes in the pandemic situation such as the number of infections, providing the possibility of describing the differences between cities' vulnerability levels. The data were from 1 March 2020, to 1 March 2022, and accessed on 8 March 2022.

For the convenience of data processing, an evaluation system was developed with Python (v3.7.9), ArcMap (v10.5), and Visual Studio Code (v1.52.1).

3.3. Pandemic Vulnerability Index

This research used FA and LambdaMART to establish PVI through a typical machine learning workflow. Sections 3.3.1–3.3.4 describe variable pre-processing, COVID-19 damage representation, LambdaMART details, and training and validation settings, respectively.

Figure 1 shows the overall framework of this research. This includes four steps:

1. Extracting influential factors related to urban pandemic vulnerability through Factor Analysis and calculating a Damage of COVID-19 Pandemic (DOP) score for the pandemic;
2. Using the urban factors and DOP score as data and labels, respectively, to supervise the training of a LambdaMART model;
3. Using the trained LambdaMART model to establish the PVI, and evaluating the PVI's performance on the validation dataset;
4. Analyzing the PVI to reveal critical factors regarding urban pandemic vulnerability.

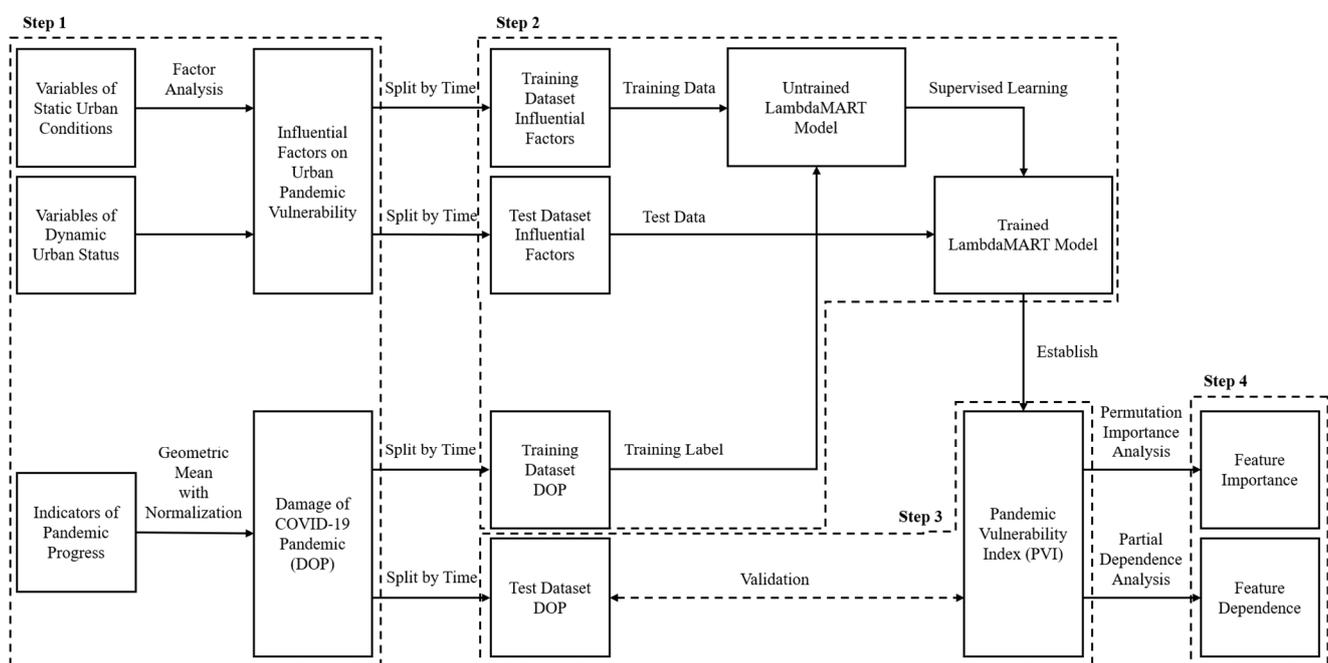


Figure 1. The Pandemic Vulnerability Index framework using machine learning.

3.3.1. Influential Variables on Vulnerability

Many variables may impact urban vulnerability, including population density, GDP, medical facilities, etc. [26–30]. Specific to pandemic vulnerability, these variables can roughly be divided into two groups: statistic variables that describe the static conditions of a city over a relatively long period, such as population, industrial structure, medical facilities; and dynamic variables that describe the dynamic status of a city during the pandemic, such as the interim policies, urban vitality, and disease prevalence [12,13,16,19].

Dynamic variables represent the efforts made to fight the pandemic, which will change over time. Although statistical variables will also change with the development of the pandemic, it is difficult to obtain the latest information due to the statistical process. Considering that changes in statistical variables are usually relatively slow, and PVI is more concerned with relative differences, statistics variables before the outbreak were used.

Since there may be a strong correlation in variables, extracting interpretable factors can effectively reduce the number of variables and facilitate subsequent calculations. A widely used method is Factor Analysis (FA), which assumes that all observed correlated variables are determined by orthogonal unobserved factors [31]. Researchers can locate a set of factors that reveal a simple hidden structure without losing the information contained in the original variables.

The FA used in this research explains a set of m variables in each of n cities with a set of k factors. There should be fewer factors than variables, so $k < m$, and these factors are related to the variable via a factor-loading matrix $L \in \mathbb{R}^{m \times k}$. The model can be concluded as follows:

$$X - M = LF + \epsilon \quad (1)$$

where observation matrix $X \in \mathbb{R}^{m \times n}$, factors matrix $F \in \mathbb{R}^{k \times n}$, error term matrix $\epsilon \in \mathbb{R}^{m \times n}$ and mean matrix $M \in \mathbb{R}^{m \times n}$. By choosing appropriate constraints, the observation matrix X can be transformed into the factor matrix F without losing too much information. The resulting factors were used together with the dynamic variables as data for subsequent model training and PVI establishment.

3.3.2. Damage of COVID-19 Pandemic

Cities' vulnerability can be represented by the damage caused by COVID-19. The greater the actual damage, the more vulnerable the city is to a pandemic. Here, we use the total score for infection status and pressures on the medical care system to present the Damage of COVID-19 Pandemic (DOP). According to the Ministry of Health, Labor, and Welfare of Japan, six indicators were officially used to characterize the COVID-19 status (see Table 1) [23].

Table 1. Status of medical care provision system in prefectures (6 indicators) Adapted from Ref. [23] from the Ministry of Health, Labor, and Welfare of Japan.

Type	Indicator	Unit
Medical care provision	Secured bed usage rate	%
	Number of recuperates	Per 100,000 people
	Positive rate in PCR test	Per 100,000 people
Monitoring system	Number of new infection cases	%
Infection status	Number of new infection cases; week-on-week ratio	-
	Unknown infection route rate	%

These indicators are all critical descriptions of the COVID-19 pandemic and represent its speed of spread, severity, and the stress on the healthcare system. A comprehensive single metric is needed to capture the damage to cities caused by the pandemic. Due to the method's simplicity and limited compensability, the geometric mean after the min–max normalization is used to aggregate these indicators [32]. A city's DOP score will be set based on the geometric mean of normalized indicators NI , and rescaled to between 0 and 10.

$$NI_{s,d} = \frac{I_{s,d} - \min I_{s,d}}{\max I_{s,d} - \min I_{s,d}}, \quad DOP_d = \sqrt[6]{\prod_s NI_{s,d}} \times 10, \quad d = \text{date} \quad (2)$$

Subscript s denotes the six different indicators, and subscript d denotes the date. The DOP score varies over time, representing the change in the damage to cities as the pandemic spreads. Note that indicators are normalized according to maximum and minimum values among cities, and a city will only receive a maximum score of 10 when all its indicators are at a maximum among cities. This normalization means that the DOP score reflects more relative disparities between cities, rather than the absolute severity of the pandemic. As the pandemic spreads exponentially, unnormalized data will show exponential shifts, obscuring the data characteristics at the beginning of the outbreak. Using normalized data will allow for a focus on relative comparisons between cities, which is in line with PVI's attempt to characterize the relative ability of cities to counter the damage caused by the pandemic.

Here, the DOP score only measures the pressure COVID-19 places on public health and does not cover subsequent damages such as economic losses or mental harm. Since research has shown that the more serious the damage to the public health system, the

more serious the economic and social damage that follows [33], the DOP score can serve as a simple, direct, and comprehensive measurement of the relative damage caused by the pandemic.

3.3.3. LambdaMART Model

A city's pandemic vulnerability depends on the impact factors F mentioned in Section 3.3.1. The factors may have different weights and influence paths, represented by a set of parameters β . The pandemic vulnerability index PVI can be written as a function of factor vector F and a set of parameters β .

$$PVI = f(F, \beta) \quad (3)$$

Traditionally, the vulnerability index is a linear function with experts assigned as β . Such a method is limited in its expressive ability and relies too much on prior knowledge regarding data differences and regional differences [34]. Here, the supervised machine learning algorithm is used to infer the f and β automatically. The machine learning algorithm will first assume the function form f and provide an initial guess of the parameters β . Then, the algorithm will compare the difference between the resulting PVI and the actual damage caused by the COVID-19 pandemic (DOP score) and update the f and β based on the differential gradient. Through iterations, the machine learning algorithm can infer suitable a f and β that best fit the empirical DOP data, which means the resulting PVI can reflect the damage and describe the urban vulnerability to the pandemic.

In this research, we chose LambdaMART due to the specificity of the PVI. Urban pandemic vulnerability is a relative concept based on inter-city comparisons. Hence, the constructed PVI should be a relative indicator that relative ranking among cities is more important than the absolute score. Fortunately, the Learning to Ranking (LTR) technique was designed to develop an optimal ordering of items and provide a ranking, which is suitable for PVI. The LambdaMART algorithm was chosen from the LTR methods due to its powerful expression ability and robustness [35].

The LambdaMART algorithm belongs to the family of decision tree algorithms, assuming the basic functional form is a decision tree [36]. For a typical decision tree, all observations x are classified into p different regions R_p , and the average of label y_p is used as the predicted value in the region:

$$y_p = \frac{1}{N_p} \sum_{x \in R_p} y_i, \quad \hat{y} = T(x) = \sum_p y_p I(x \in R_p) \quad (4)$$

Usually, a single decision tree will not produce a good prediction result. The Multiple Additive Regression Tree (MART) will iteratively calculate the loss between observed DOP score and predicted PVI, fit new decision trees along the differential gradient of previous prediction loss, and the final result will be the sum of all decision trees. However, here the DOP score and PVI represent a relative ranking, which makes it challenging to compute a differentiable loss. Therefore, LambdaMART uses a pairwise method to transform the DOP score into a partial order of pairwise comparisons. For city i and city j , the actual probability of city i being more vulnerable than city j is denoted as P_{ij} :

$$P_{ij} = \frac{1}{2}(1 + S_{ij}), \quad S_{ij} = \text{Sgn}(DOP_j - DOP_i) \quad (5)$$

While the possibility given by the LambdaMART model is \hat{P}_{ij} :

$$PVI_i = M(F_i), \quad PVI_j = M(F_j), \quad \hat{P}_{ij} \equiv P(PVI_i > PVI_j) \equiv \frac{1}{1 + e^{-\sigma(PVI_i - PVI_j)}} \quad (6)$$

Therefore, the loss between observed and predicted can take the differentiable cross-entropy form:

$$C \equiv -P_{ij} \log \hat{P}_{ij} - (1 - P_{ij}) \log(1 - \hat{P}_{ij}) = \frac{1}{2}(1 - S_{ij})\sigma(PVI_i - PVI_j) + \log\left(1 + e^{-\sigma(PVI_i - PVI_j)}\right) \quad (7)$$

It should be noted that the loss function here treats all cities equally. However, we are more concerned with those cities that are ranked higher and are more vulnerable. LambdaMART introduced the Normalized Discounted Cumulative Gain (NDCG), which emphasizes samples with high rankings. Therefore, gradient λ can be defined on the partial derivative of loss C and NDCG measurements.

$$\lambda_{ij} \equiv \frac{\partial C}{\partial PVI_i} \times |\Delta NDCG| = \sigma\left(\frac{1}{2}(1 - S_{ij}) - \frac{1}{1 + e^{-\sigma(PVI_i - PVI_j)}}\right) \times |\Delta NDCG| \quad (8)$$

where the $|\Delta NDCG|$ represents the difference in NDCG after exchanging the positions of i and j . A new decision tree T^{l+1} now can be fit on gradient λ^l from the latest decision tree T^l . After L iterations, the PVI given by the LambdaMART algorithm will be as follows:

$$PVI = M(F, \beta) = \sum_L T^L(x) \quad (9)$$

In short, the LambdaMART model will repeat the cycle of “fitting a decision tree—obtaining PVI—measuring the difference between PVI and DOP scores—calculating gradient—fitting a new decision tree” until the difference between the observed DOP score and predicted PVI is small enough in terms of the partial order of pairwise comparisons.

3.3.4. Training and Validation

For machine learning, overfitting is a critical problem, which means that the LambdaMART model pays too much attention to the existing data and loses the ability to work on unobserved data. In the context of this research, overfitting means that the established PVI is consistent with the observed DOP score but cannot make a valid prediction for the future.

A general solution is the train-test splitting technique. The dataset is divided into two parts, the training dataset and the test dataset, and the model is trained using only the training dataset. When the model achieves an excellent performance and the established PVI is consistent with the observed DOP score, the model is then validated on the test dataset to see if the resulting PVI reflects the “unobserved” DOP score.

This train-test splitting technique can help us evaluate how accurately the established PVI measures urban pandemic vulnerability. In this research, the two-year dataset was evenly split into training and test datasets according to time. The data from 1 March 2020 to 1 March 2021 formed the training set used to train the LambdaMART model. The data from 1 March 2021 to 1 March 2022 formed the test dataset, used to verify the model’s performance. The data from the Diamond Princess cruise ship and imported cases were omitted to focus on vulnerability in the urban area.

Another time-related problem is the lag in pandemic damage. At a given moment, the pandemic vulnerability will not be immediately reflected in the DOP score at that exact moment, but instead will be delayed for a while. Considering that the incubation period of COVID-19 can extend up to 24 days, the PVI of cities at a specific moment should be able to predict the DOP score of the next period. According to Lauer, et al. [37], 99% of patients will develop symptoms in 14 days. Therefore, the time lag for the DOP score is set to 14 days, which means that PVI at day d is used to characterize the pandemic damage at day $d + 14$.

4. Results

4.1. Variables Selection

Aiming to explore the possible relevant variables that affect urban pandemic vulnerability, this research referred to the variables included in previous research [12,13,16,19]. This

research used 140 variables from the Digital National Land Database, 6 variables from the Google Community Mobility Report, and 10 variables from the Ministry of Health, Labor, and Welfare. The complete variable list is given in Table S1 in the Supplementary Materials.

These variables include statistic variables that describe the static conditions over a period and dynamic variables that change over time. The statistic variables involved in this research can be divided into the following five aspects:

- **Demographic Variables.** Intuitively, the scale of a city is closely related to the spread of infectious diseases, and overpopulated cities are more vulnerable to a pandemic. Variables such as urban built-up area population density are included.
- **Economic Variables.** Active economic activity means that more urban resources can be mobilized to counter pandemics, and diseases are more easily spread. Fiscal expenditures closely related to economic activities contribute to improved medical and public facilities.
- **Mobility indicators.** Population movement between different regions provides conditions for the pandemic, including the inflow and outflow of the population, the proportion of the daytime population, etc.
- **Spatial Variables.** The spatial structure of different cities is the most important factor that constitutes the difference in urban internal spatial activities. The proportions of various land-use types are included.
- **Medical Variables.** Medical and health services will also affect the spread of infectious diseases in cities. This includes the number of service facilities, the number of medical practitioners and related financial expenditures, etc.

Figure 2 visualizes some statistical variables, and the differences between regions can be seen. The number of employees and commuters is mainly concentrated around large cities, while the number of beds, as essential medical resources, is relatively evenly distributed.

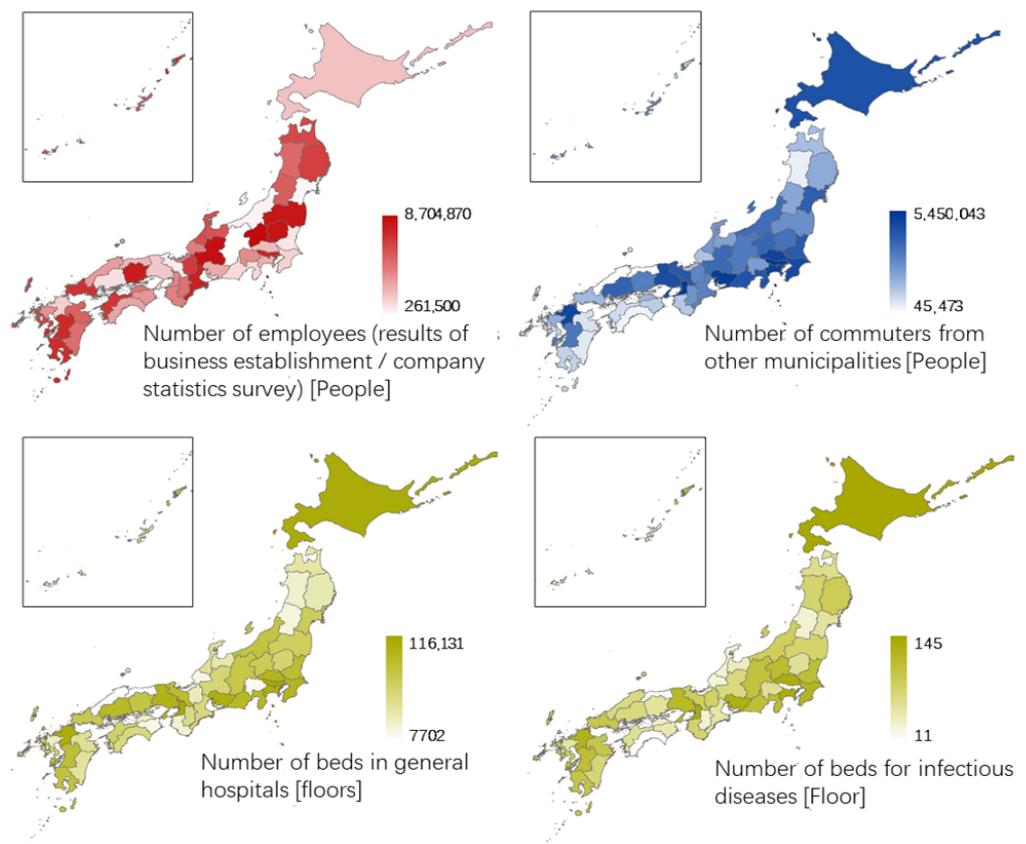


Figure 2. Visualization of some statistical variables.

The included dynamic variables can be divided into the following three aspects:

- **Vitality Variables.** The number of people active in different urban areas is compared with the baseline value of February 2020, which can help characterize the urban vitality changes that reflect residents' reactions to the pandemic. Urban functioning areas is classified into six types: retail and recreation, grocery and pharmacy, parks, transit stations, workplaces, and residential areas.
- **Policy Variables.** Whether a prefecture announces Emergency Status (Japanese: 緊急事態宣言) or takes Key Measures for Spread Prevention (Japanese: まん延防止等重点措置) are used as binary variables to describe the policy reaction. Whether the day is a holiday or weekend is also included.
- **Pandemic variables.** The six indicators mentioned in Section 3.3.2 are included when calculating the DOP score.

Figure 3 shows the daily new COVID-19 infections per 100,000 people and the urban vitality in Tokyo, Osaka, and Ishikawa. The red bars represent daily new COVID-19 infections on the log scale, showing six pandemic waves. The solid lines represent the vitality of urban functional areas compared to before the pandemic: residential (red), retail (blue), transit (yellow), and workplace (green). The lower horizontal bar represents the status of the city. The red means the city declared an Emergency Status in that period, the orange means the city took Key Measures for Spread Prevention, and the black line means that this was a nationwide holiday.

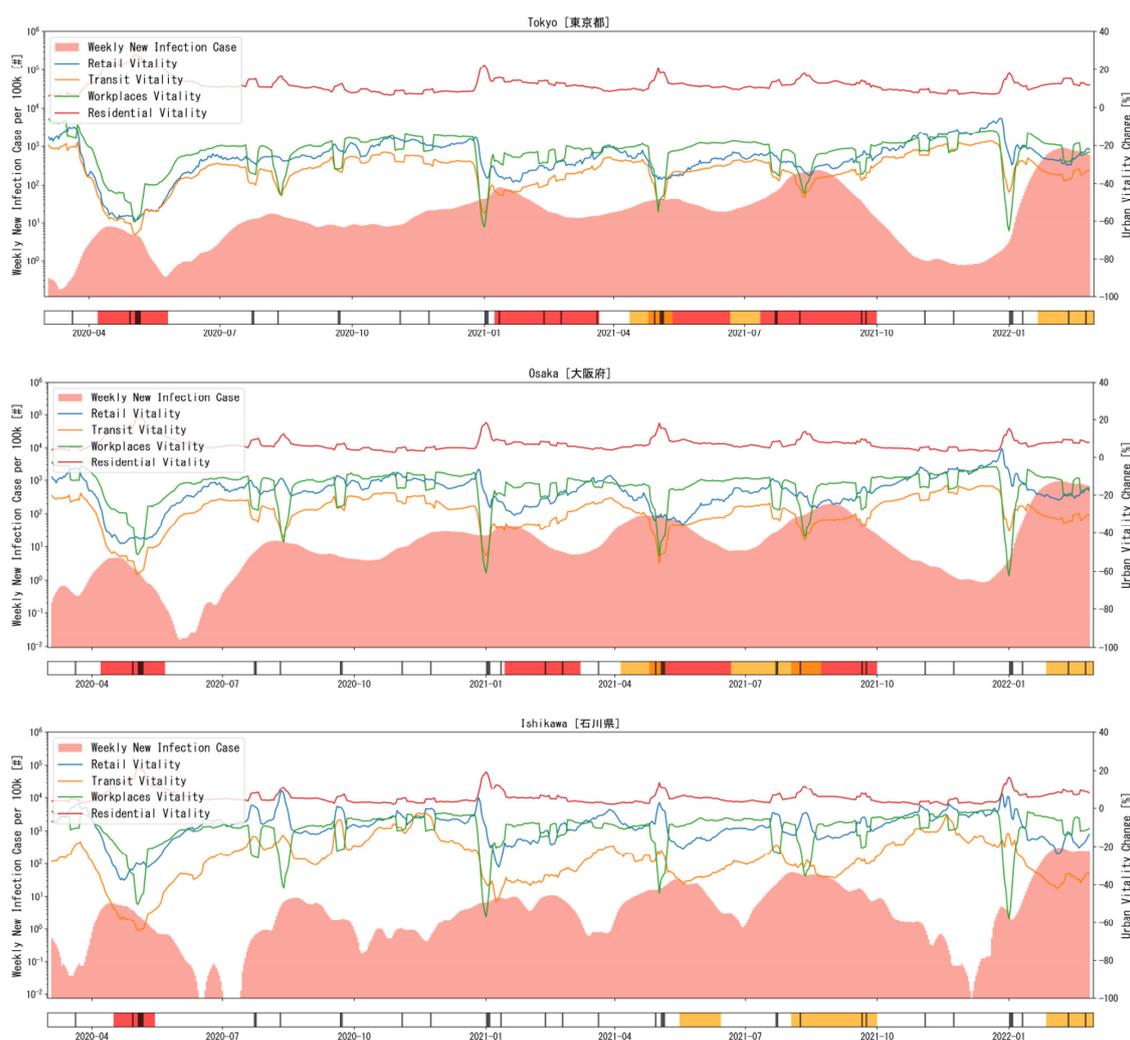


Figure 3. Daily COVID-19 cases per 100,000 people and Urban Vitality from 1 March 2020 to 1 March 2022 in Tokyo (upper), Osaka (middle), and Ishikawa (lower).

It can be seen from Figure 3 that urban life gradually returned to a new balance after the initial shock of the pandemic in early 2020, with the apparent fluctuations all being holiday-related. The urban vitality in the residential area increased by about 15% compared to before the pandemic, which may be related to the work-from-home trend. The urban vitality of workplaces, transit, and retail significantly decreased, with Tokyo down by about 30%, Osaka by 20%, and Ishikawa by 10%, showing that the impact of the pandemic varied depending on city size. The urban vitality of the transit area was comparatively the most affected, followed by the retail area. On the other hand, the pandemic waves appear to be linked to holidays and the associated urban vitality changes. In all three areas, the declaration of an emergency status and measures for spread prevention seem to be helpful to control the pandemic.

4.2. Factor Analysis

The Factor Analysis method mentioned in Section 3.3.1 was used to extract factors due to the strong correlation between the statistical variables.

The oblimin rotation method was adopted, and 85% of the variance was retained. Finally, nine factors were selected to characterize these 140 indicators, retaining 86.8% variation. Figure 4 shows the correlation matrix after the oblimin rotation with a correlation between factors lower than 0.32, which means that there is less than a 10% overlap in variance among factors [38]. The complete loading matrix is shown in Figure S1 in the Supplementary Materials.



Figure 4. Correlation matrix between factors after oblimin rotation.

For these nine factors, the variable with the most considerable load was extracted. This can be named according to the direction of its main load concentration (see Table 2). According to their eigenvalues, these factors were named city size, medical facilities, age structure, unemployment, cultural facilities, precipitation, industry, decentralization, and commerce. These factors and dynamic variables constitute the data in the subsequent LambdaMART model.

Table 2. The most extensive load description of factors and the name of each factor.

Factor	Describe	Name
0	Positive load on the number of households and population density	City Size
1	Positive load on the number of medical facilities and medical staff	Medical Facilities
2	Positive load of the proportion of the adolescent population Negative load of death	Age Structure
3	Positive load of the complete unemployment rate Negative load of employment rate	Unemployment
4	Positive load of the number of cultural facilities such as museums	Cultural Facilities
5	Negative load of sunshine time, Positive load of annual precipitation	Precipitation
6	Negative load of population exodus Positive load of industrial land ration	Industry
7	Negative load of population ratio in densely populated areas	Decentralization
8	Negative load of the commercial land ratio	Commerce

Figure 5 reveals the difference in several factors in Japan. Large cities are concentrated around Tokyo, Osaka, and Nagoya, metropolitan areas, while the medical facility factor is highest in South Tohoku and South Kyushu. There is a relatively large aging population in Hokkaido and the Tohoku region, and cultural facilities are concentrated in the Kinki region. These differences represent the differences in regions and may affect the urban pandemic vulnerability.

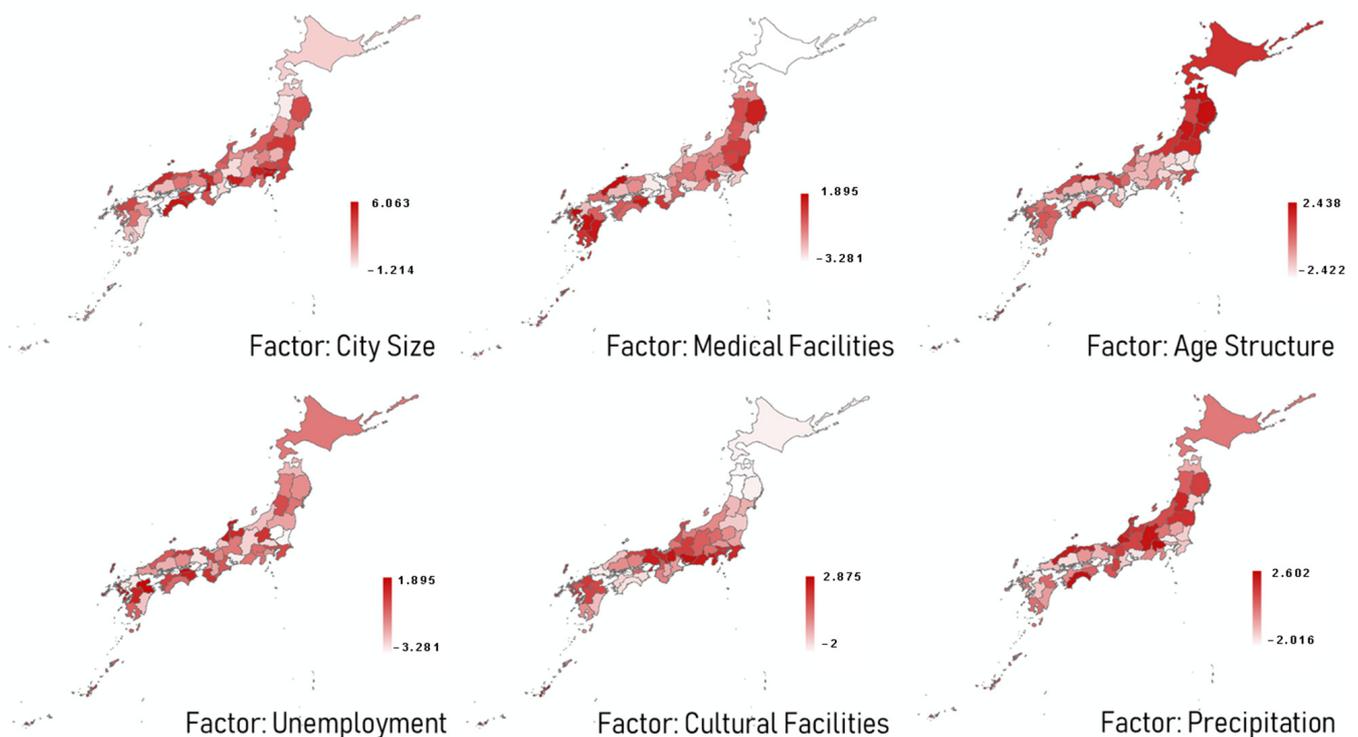


Figure 5. Visualization of some factors.

4.3. Model Performance

The LambdaMART model described in Section 3.3.4 was implemented in Python with the LightGBM package, and its hyperparameters were set as shown in Table 3.

Table 3. Hyperparameters of LambdaMART Model.

Hyperparameters	Value
Boosting type	GBDT
Number of leaves	15
Learning rate	0.05
N estimators	100
Subsample	0.8

After training, the PVI established by the LambdaMART model can be matched with the DOP score in the training dataset.

Figure 6 shows two example results in the training set. The red bars show the actual DOP score after 14 days, while the blue bars show the PVI that was learned in the training set. In the first pandemic wave on 16 May 2020, the three most vulnerable regions were Fukuoka, Ishikawa, and Hokkaido, with the highest DOP score obtained after 14 days. Similarly, on 6 February 2021, in the third pandemic wave, the three areas with the highest PVI became Tokyo, Chiba, and Kanagawa as the situation changed, and the DOP score after 14 days also changed. Noticed that the accuracy is lower for regions with a lower PVI, which is related to the NDCG metric used in the LambdaMART model. The NDCG metric assigns a higher weight to the top-ranked predictions limits model performance in less vulnerable area predictions.

Overall, the PVI and the DOP score fit perfectly, with an average $NDCG@10 = 0.9411$ and $Precision@10 = 0.7591$ in the entire training dataset. The PVI ranking is validated in the test dataset based on the learned model.

Figure 7 shows the two example results in the test set on 15 May 2021, and 8 January 2022. The DOP score in the test dataset is “unseen” for the LambdaMART model, so the result represents the model’s predictive ability. On 15 May 2021, the three regions with the highest PVI reported by the model were Okinawa, Osaka, and Hokkaido. The actual DOP score obtained 14 days later shows Okinawa, Hokkaido, and Osaka, the same regions, with slightly different rankings. The results for 8 January 2022 also show correct predictions but slightly different rankings, with only one wrong prediction in the top-10 PVI areas. Overall, the model is accurate, reporting an average $NDCG = 0.9149$ and $Precision@10 = 0.7189$ in the whole test dataset. Despite the slight drop, the model can still effectively reflect the severity of the pandemic.

Figure 8 takes three typical regions, namely Tokyo, Osaka, and Ishikawa, to represent the results in both datasets. The red line shows the actual DOP scores with a 14 day lag. The solid blue line shows the PVI in the training phase, and the blue dashed line shows the prediction of PVI. In Tokyo, the PVI effectively reflects the damage caused by COVID-19 predictions, except from October 2021 to December 2021. The calculated PVI appears to overestimate the vulnerability of Tokyo during this period. Osaka’s PVI performance is generally excellent, with occasional deviations in the test dataset. In Ishikawa, the predicted PVI appears to underestimate vulnerability between October and December 2021.

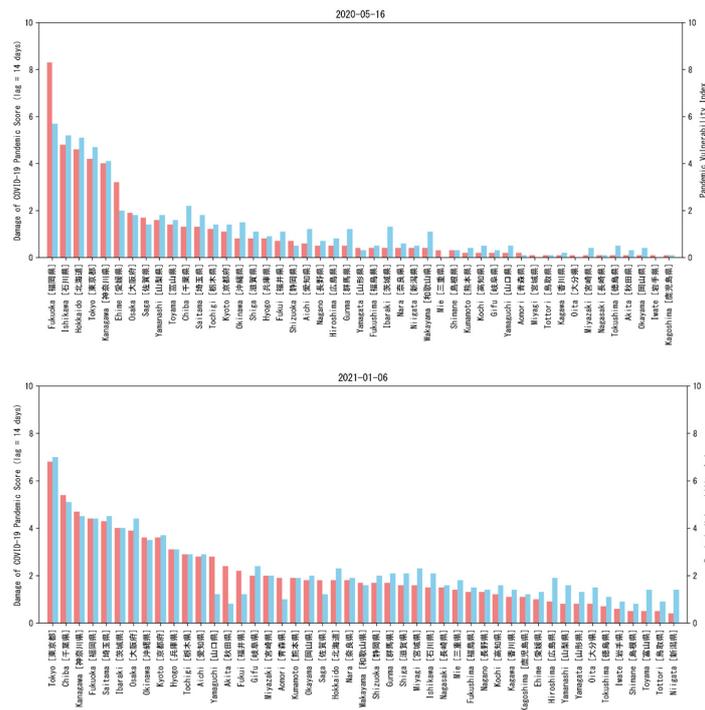


Figure 6. The DOP Score and PVI in each prefecture on 16 May 2020 (upper) and on 6 January 2021 (lower).

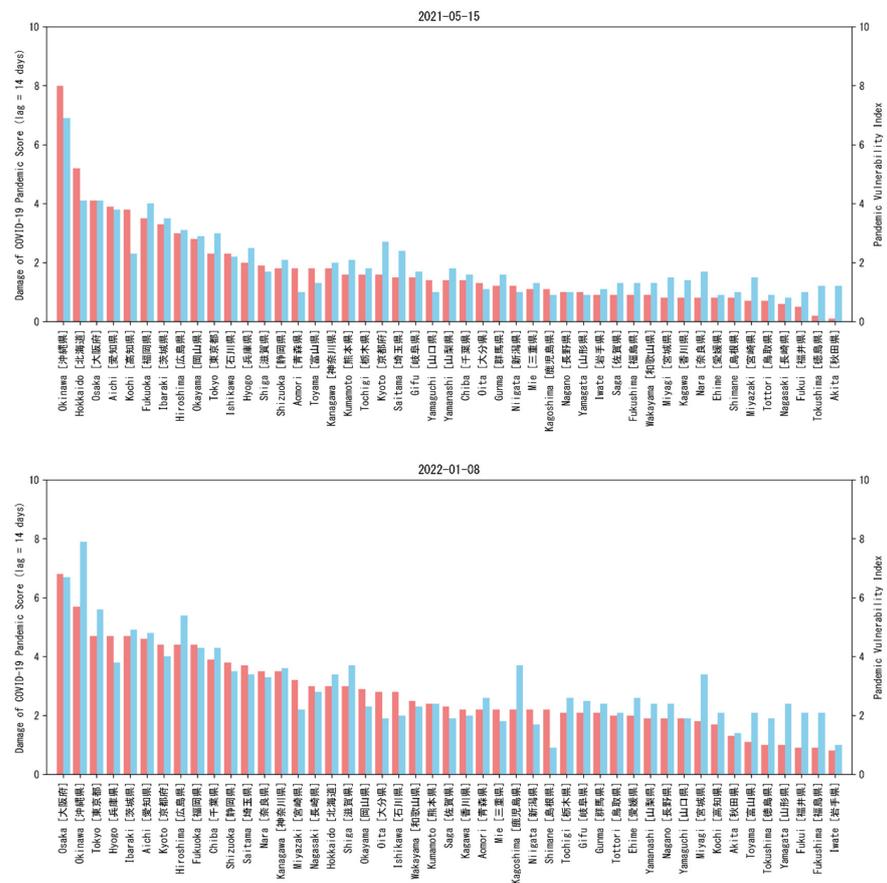


Figure 7. The DOP Score and PVI in each prefecture on 15 May 2021 (upper) and 8 January 2022 (lower).

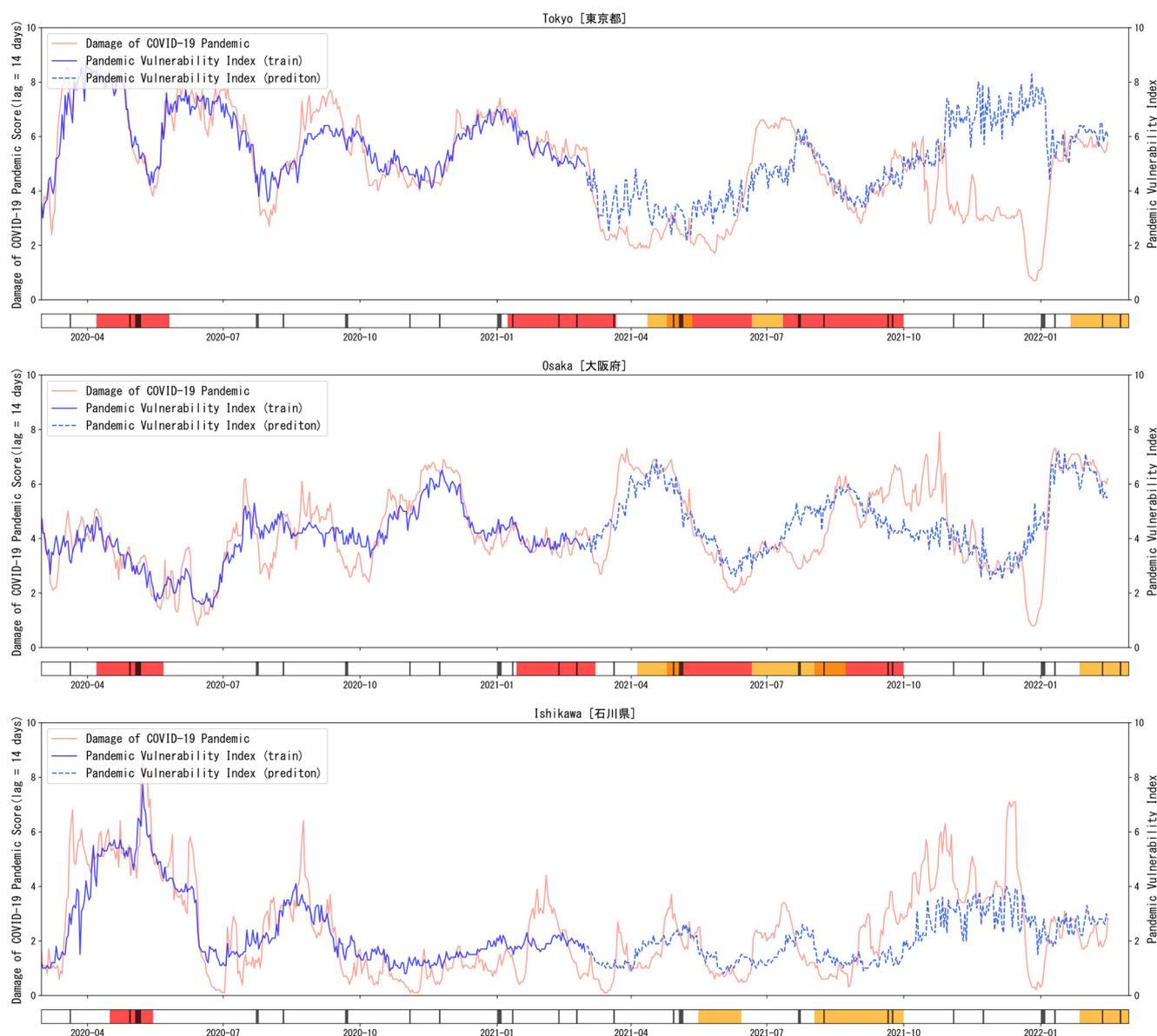


Figure 8. The COVID-19 Score and PVI in Tokyo (upper), Osaka (middle), and Ishikawa (lower) in the whole dataset.

Since PVI is a relative indicator, the cessation of the pandemic in metropolitan areas between October and December 2021 makes the model simultaneously “overestimate” the risk in the metropolitan area and “underestimate” the risk in the non-metropolitan area. However, the rapid spread of the Omicron variant in Tokyo in January 2022 (see Figure 3) shows that such a deviation is only temporary, and the metropolitan area is still vulnerable to a pandemic.

Generally, these results show that the LambdaMART model has a good generalization ability, proving that the PVI can effectively predict the damage caused by the COVID-19 pandemic in the city with an overall 0.7198 top-10 accuracy. Since PVI can forecast vulnerable regions in two weeks given real-time data, the possibility of preventive measures in advance is opened up. Such measures include, but are not limited to, social distancing, supporting healthcare needs, expanding healthcare facilities, and framing strategies to mitigate the infection. Society can also reorganize smoothly without sudden changes by managing inventory, facilitating working from home, and preparing supplies [39].

Therefore, the proposed PVI is meaningful in controlling the pandemic and shaping the response in advance.

4.4. Feature Importance and Dependence

It would be more helpful to look into the features that influenced PVI, which can guide the subsequent policy formulation and urban planning process. In this research, the Permutation Importance Analysis and Partial Dependence Analysis were carried out to examine further the obtained model.

The permutation importance analysis can evaluate each feature's importance by randomly shuffling a single feature value [40]. Figure 9 shows the permutation feature importance of PVI, where the city size and the vitality of transit stations have the highest permutation importance, about 0.31 and 0.22, respectively. A metropolitan area's city-scale and population density create sufficient conditions for a pandemic. At the same time, the dynamism of a transit station is somewhat representative of whether there is rapid population mobility and is also essential for assessing a region's urban pandemic vulnerability.

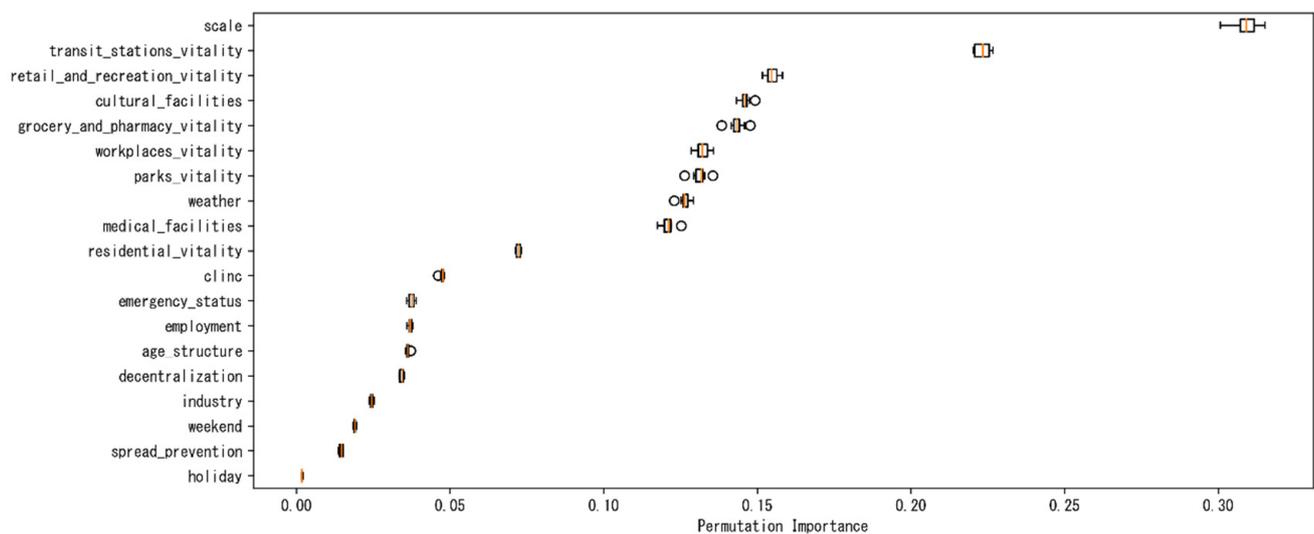


Figure 9. The permutation feature importance of PVI.

It can be seen that most dynamic variables have relatively high levels of importance, which confirms the view that it is difficult to assess urban pandemic vulnerability by relying only on static statistics. In addition to city size, the critical static factors include cultural facilities, weather, and medical facilities. Note that neither emergency status nor key measures of spread prevention seem to be important, possibly because these tend to be remedial measures, while the PVI emphasizes vulnerability before the pandemic hits.

Figure 10 further reveals the partial dependence of the critical features of PVI. The partial dependence is the expected response as a function of the input features, assuming other conditions remain unchanged, shown as the thick blue line. The light blue line indicates the individual conditional expectation separately, with one line per sample. These features show different patterns. The increase in city-scale and transit station vitality will lead to an increase in PVI, showing that cities with large populations and high mobility will have high pandemic vulnerability. The increase in parks' vitality and medical facilities leads to a decrease in pandemic vulnerability.

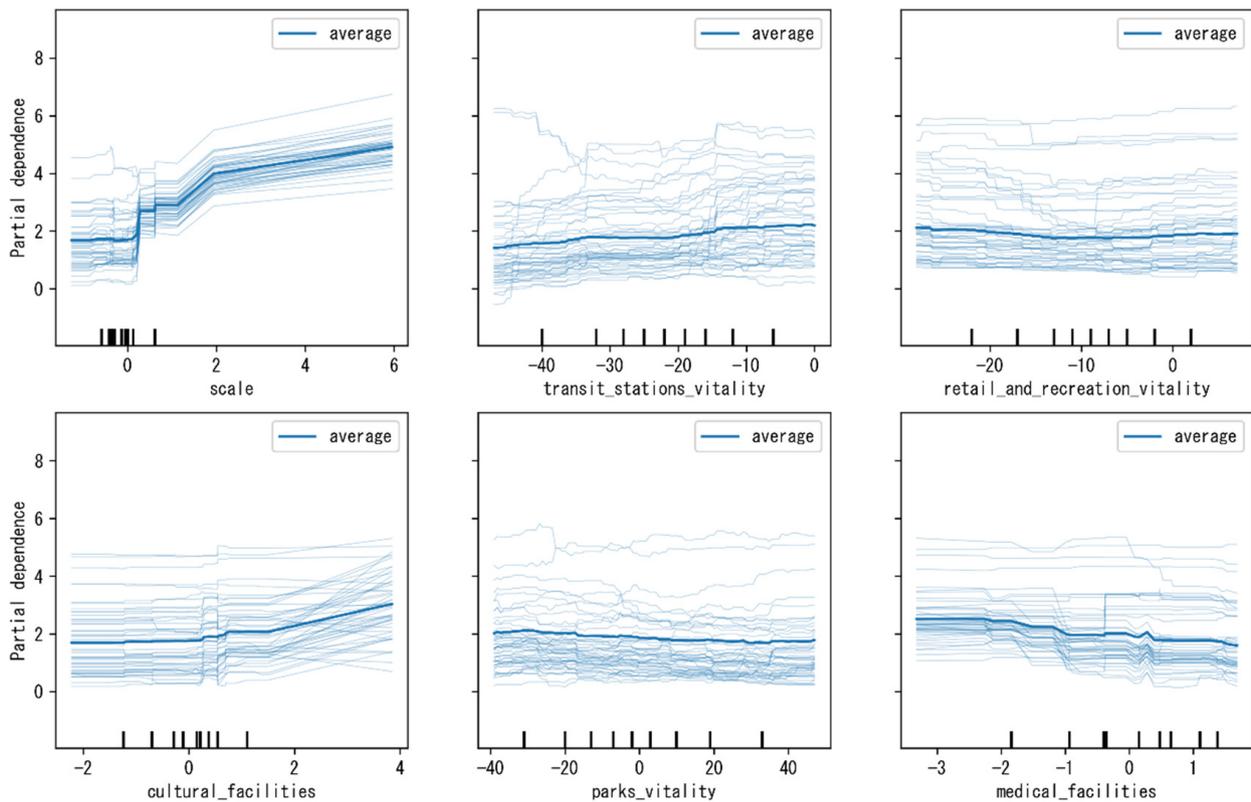


Figure 10. The partial dependence of the top six features in the PVI.

As Figures 9 and 10 indicate, city-scale and transit station vitality are the two most important factors. The city's expansion has a relationship with the city's ability to resist the risk of infectious diseases and increase the city's vulnerability. For Japan, the three major metropolitan areas are at the core of social and economic development, meaning that metropolitan pandemic risk will be an essential issue for future planning. The vitality of the transit station is both a factor affecting urban vulnerability and a target of pandemic impact, making causality more challenging to analyze. However, in any case, the public transport system will be a weak link for cities when facing the epidemic. In addition, the PVI drops when medical facilities feature increases, showing that medical infrastructure investment might provide advantages in fighting COVID-19.

Changes in urban vitality also have significant impacts. The rise of urban vitality in transit leads to an increase in PVI, which verifies the necessity of the social distancing policy. On the other hand, the park's vitality helps urban vulnerability, suggesting that public open spaces such as parks should attract more attention from urban planners in the post-COVID era.

5. Conclusions and Discussion

COVID-19 has completely changed urban life and brought new problems and challenges to urban vulnerability research. In response to these challenges, this research proposed the concept of urban pandemic vulnerability as the first step to supplementing the urban vulnerability research and providing a Pandemic Vulnerability Index, using Japan as an example.

In this research, we took a series of statistic variables and dynamic variables of the city as a base, used the Factor Analysis to reduce the dimension, calculated the Damage of COVID-19 Pandemic Score to evaluate the damage caused by the pandemic, and used the LambdaMART algorithm to establish a Pandemic Vulnerability Index that targets the critical characteristics regarding pandemic vulnerability. The results indicate that the PVI

proposed can effectively predict the damage caused by the COVID-19 pandemic, and further analysis revealed the key features that should be focused on to reduce pandemic vulnerability. This method could be applied to flexible data and different regions.

The main contributions of this research are:

- This research established a Pandemic Vulnerability Index that can indicate relative urban vulnerability and incorporate dynamic factors into indicator construction.
- LambdaMART is efficient in constructing a relative ranking index for urban vulnerability and can predict infection development with high precision. Accurate short-term forecasts help to take advance measures and help with preparation.
- Feature importance and dependence analysis emphasize city-scale and transit station vitality when evaluating urban pandemic vulnerability.

Compared with related studies, this research has made significant improvements. The Urban Vulnerability Assessment proposed by Prieto, Malagón, Gomez and León [12] combines information on demographic factors, work styles, and transportation through Borda Counting. However, the variables used in [12] come from surveys taken before the pandemic, and the method is designed to respond to static geographic data; therefore, it cannot reflect urban dynamics during the development of the pandemic and is struggles to guide timely action through analysis. Our research has included dynamic factors such as urban vitality in addition static data, demonstrated dynamic factors' importance in evaluating urban pandemic vulnerability, and provided a reference for preventive measures in advance by forecasting two weeks in advance. Jardim, Castro Neto, Alpalhão and Calçada [16] presented an Urban Dynamic Indicator through time series decomposition and factor analysis. However, the proposed indicator aims to provide an alternative reference for urban vitality and cannot be directly applied to vulnerability assessments of the COVID-19 pandemic. The PVI proposed in our research is also a dynamic indicator and can reflect the damage to the city in two weeks, which is beyond simple descriptions of urban vitality. Overall, this research responds to the limited advances in urban dynamics in urban vulnerability assessments [10] and improves the capacity to capture the dynamic nature of urban vulnerability.

There are still some limitations to this research. Due to a lack of data, some important indicators, such as vaccination status, are not covered. Although the method does not depend on specific indicators, the lack of essential indicators may impact the model's performance and interpretation. Although PCA can retain critical information from original variables, the PVI's robustness to variable selection still needs further study. Additionally, this research is based on COVID-19 target data, which puts forward higher requirements for the data collection process in developing countries. When discussing inequality issues in small regions, such as a city block, such data requirements may pose certain obstacles, while looking into the vulnerable regions in the city is critical to deepening the understanding of urban vulnerability. Although a DOP Score is proposed as a reference, the damage to cities caused by the COVID-19 pandemic is complex, comprehensive, and has not been fully assessed to date. In the absence of a rational justification for assigning weights, which needs to be developed in future research, the proposed DOP scores assumed equal importance for all involved indicators. The relative importance of variables in building such a composite indicator calls for further in-depth analysis [12,41].

The analysis shows that the public transportation system is the weak link in the city's response to the pandemic, so we recommend paying attention to the density of the public transportation system when facing a pandemic and encouraging response measures such as wearing masks. Excessive city size can also make cities more vulnerable to pandemics, so we suggest that new urban planning ideas such as compact cities should be examined carefully in future urban planning. COVID-19 will permanently change our world, but a healthy, livable city will be the constant pursuit. The development of new technology will be our continuous progress in dealing with urban vulnerability and proposing proper urban planning based on facts.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/su14095053/s1>, Figure S1: Loading Matrix of Factor Analysis; Table S1: Statistic and Dynamic Variables List.

Author Contributions: Conceptualization, Y.L. and Z.S.; data curation, Y.L.; formal analysis, Y.L.; methodology, Y.L.; software, Y.L.; supervision, Z.S.; validation, Y.L. and Z.S.; visualization, Y.L.; writing—original draft, Y.L.; writing—review and editing, Y.L. and Z.S. 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: Publicly available datasets were analyzed in this study. These data can be found here: [17–22].

Acknowledgments: Special thanks to Yajing Zhang for her advice and support in this research.

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

References

- Sharifi, A. The COVID-19 Pandemic: Lessons for Urban Resilience. In *COVID-19: Systemic Risk and Resilience*; Linkov, I., Keenan, J.M., Trump, B.D., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 285–297.
- Romero Lankao, P.; Qin, H. Conceptualizing urban vulnerability to global climate and environmental change. *Curr. Opin. Environ. Sustain.* **2011**, *3*, 142–149. [[CrossRef](#)]
- Mishra, S.V.; Gayen, A.; Haque, S.M. COVID-19 and urban vulnerability in India. *Habitat Int.* **2020**, *103*, 102230. [[CrossRef](#)] [[PubMed](#)]
- Enright, T.; Ward, K. Governing urban infrastructures under pandemic conditions: Some thoughts. *Urban Geogr.* **2021**, *42*, 1023–1032. [[CrossRef](#)]
- Rufat, S. Spectroscopy of Urban Vulnerability. *Ann. Assoc. Am. Geogr.* **2013**, *103*, 505–525. [[CrossRef](#)]
- Rasch, R.J. Assessing urban vulnerability to flood hazard in Brazilian municipalities. *Environ. Urban.* **2015**, *28*, 145–168. [[CrossRef](#)]
- Hoekman, L.M.; Smits, M.M.V.; Koolman, X. The Dutch COVID-19 approach: Regional differences in a small country. *Health Policy Technol.* **2020**, *9*, 613–622. [[CrossRef](#)]
- Azzolina, D.; Lorenzoni, G.; Silvestri, L.; Prosepe, I.; Berchiolla, P.; Gregori, D. Regional Differences in Mortality Rates During the COVID-19 Epidemic in Italy. *Disaster Med. Public Health Prep.* **2020**, 1–7. [[CrossRef](#)]
- UN Habitat. *COVID-19 Response Report*; UN Habitat: Nairobi, Kenya, 2020.
- Salas, J.; Yepes, V. Urban vulnerability assessment: Advances from the strategic planning outlook. *J. Clean. Prod.* **2018**, *179*, 544–558. [[CrossRef](#)]
- Rigillo, M.; Cervelli, E. Mapping Urban Vulnerability: The Case Study of Gran Santo Domingo, Dominican Republic. *Adv. Eng. Forum* **2014**, *11*, 142–148. [[CrossRef](#)]
- Prieto, J.; Malagón, R.; Gomez, J.; León, E. Urban Vulnerability Assessment for Pandemic Surveillance—The COVID-19 Case in Bogotá, Colombia. *Sustainability* **2021**, *13*, 3402. [[CrossRef](#)]
- Shi, C.; Liao, L.; Li, H.; Su, Z. Which urban communities are susceptible to COVID-19? An empirical study through the lens of community resilience. *BMC Public Health* **2022**, *22*, 70. [[CrossRef](#)] [[PubMed](#)]
- Sharifi, A.; Khavarian-Garmsir, A.R. The COVID-19 pandemic: Impacts on cities and major lessons for urban planning, design, and management. *Sci. Total Environ.* **2020**, *749*, 142391. [[CrossRef](#)] [[PubMed](#)]
- Sharma, A.; Borah, S.B. COVID-19 and Domestic Violence: An Indirect Path to Social and Economic Crisis. *J. Fam. Violence* **2020**. [[CrossRef](#)] [[PubMed](#)]
- Jardim, B.; Castro Neto, M.d.; Alpalhão, N.; Calçada, P. The daily urban dynamic indicator: Gauging the urban dynamic in Porto during the COVID-19 pandemic. *Sustain. Cities Soc.* **2022**, *79*, 103714. [[CrossRef](#)]
- TV-Asahi. The Number of Newly Infected People Exceeds 4500 Nationwide, Which Is the Highest in Tokyo and 5 Other Prefectures. 2021.01.01. 2021. Available online: https://news.tv-asahi.co.jp/news_society/articles/000202838.html (accessed on 22 March 2022).
- Okinawa Times. Okinawa Infection “the Most Experienced in Japan in the Past” The Maximum Number of Recuperators Was 2208. 2021.05.28. 2021. Available online: <https://www.okinawatimes.co.jp/articles/-/760946> (accessed on 22 March 2022).
- Zawbaa, H.M.; El-Gendy, A.; Saeed, H.; Osama, H.; Ali, A.M.A.; Gomaa, D.; Abdelrahman, M.; Harb, H.S.; Madney, Y.M.; Abdelrahim, M.E.A. A study of the possible factors affecting COVID-19 spread, severity and mortality and the effect of social distancing on these factors: Machine learning forecasting model. *Int. J. Clin. Pract.* **2021**, *75*, e14116. [[CrossRef](#)]
- Pan, Y.; Zhang, L.; Yan, Z.; Lwin, M.O.; Skibniewski, M.J. Discovering optimal strategies for mitigating COVID-19 spread using machine learning: Experience from Asia. *Sustain. Cities Soc.* **2021**, *75*, 103254. [[CrossRef](#)]

21. Ministry of Land Infrastructure Transport and Tourism. National Land Numerical Information. Available online: <https://nlftp.mlit.go.jp/ksj/> (accessed on 7 March 2022).
22. Ministry of Health Labour and Welfare. COVID-19 Trends & Current Situation. Available online: https://www.mhlw.go.jp/stf/covid-19/kokunainohasseijoukyou_00006.html (accessed on 7 March 2022).
23. Ministry of Health Labour and Welfare. Indicators to Assess the Level of Community Transmission. Available online: <https://covid19.mhlw.go.jp/extensions/public/en/index2.html> (accessed on 7 March 2022).
24. Google. Community Mobility Reports. Available online: <https://www.google.com/covid19/mobility/> (accessed on 7 March 2022).
25. ESRI. Coronavirus Support Site. Available online: <https://coronavirus-esri-japan-ej.hub.arcgis.com/> (accessed on 7 March 2022).
26. Takano, T.; Nakamura, K. An analysis of health levels and various indicators of urban environments for Healthy Cities projects. *J. Epidemiol. Community Health* **2001**, *55*, 263–270. [[CrossRef](#)]
27. Ison, E. The introduction of health impact assessment in the WHO European Healthy Cities Network. *Health Promot. Int.* **2009**, *24*, i64–i71. [[CrossRef](#)]
28. Webster, P.; Sanderson, D. Healthy Cities indicators—A suitable instrument to measure health? *J. Urban Health* **2013**, *90*, 52–61. [[CrossRef](#)]
29. Pineo, H.; Glonti, K.; Rutter, H.; Zimmermann, N.; Wilkinson, P.; Davies, M. Urban Health Indicator Tools of the Physical Environment: A Systematic Review. *J. Urban Health* **2018**, *95*, 613–646. [[CrossRef](#)]
30. Hu, M.; Roberts, J.D.; Azevedo, G.P.; Milner, D. The role of built and social environmental factors in Covid-19 transmission: A look at America’s capital city. *Sustain. Cities Soc.* **2021**, *65*, 102580. [[CrossRef](#)]
31. Harman, H.H. *Modern Factor Analysis*; University of Chicago Press: Chicago, IL, USA, 1976.
32. Greco, S.; Ishizaka, A.; Tasiou, M.; Torrissi, G. On the Methodological Framework of Composite Indices: A Review of the Issues of Weighting, Aggregation, and Robustness. *Soc. Indic. Res.* **2018**, *141*, 61–94. [[CrossRef](#)]
33. Boettke, P.; Powell, B. The political economy of the COVID-19 pandemic. *South. Econ. J.* **2021**, *87*, 1090–1106. [[CrossRef](#)] [[PubMed](#)]
34. Rothenberg, R.; Weaver, S.R.; Dai, D.; Stauber, C.; Prasad, A.; Kano, M. A flexible urban health index for small area disparities. *J. Urban Health* **2014**, *91*, 823–835. [[CrossRef](#)]
35. Burges, C.J. From ranknet to lambdarank to lambdamart: An overview. *Learning* **2010**, *11*, 81.
36. Prasad, A.M.; Iverson, L.R.; Liaw, A. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems* **2006**, *9*, 181–199. [[CrossRef](#)]
37. Lauer, S.A.; Grantz, K.H.; Bi, Q.; Jones, F.K.; Zheng, Q.; Meredith, H.R.; Azman, A.S.; Reich, N.G.; Lessler, J. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann. Intern. Med.* **2020**, *172*, 577–582. [[CrossRef](#)]
38. Corner, S. Choosing the right type of rotation in PCA and EFA. *JALT Test. Eval. SIG Newsl.* **2009**, *13*, 20–25.
39. Devaraj, J.; Madurai Elavarasan, R.; Pugazhendhi, R.; Shafiullah, G.M.; Ganesan, S.; Jeysree, A.K.; Khan, I.A.; Hossain, E. Forecasting of COVID-19 cases using deep learning models: Is it reliable and practically significant? *Results Phys.* **2021**, *21*, 103817. [[CrossRef](#)]
40. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
41. Salas, J.; Yepes, V. VisualUVAM: A Decision Support System Addressing the Curse of Dimensionality for the Multi-Scale Assessment of Urban Vulnerability in Spain. *Sustainability* **2019**, *11*, 2191. [[CrossRef](#)]