ISSN 2616-5872 Vol.7, Issue 3: 55-67, DOI: 10.25236/AJEE.2025.070308

Research on the Resilience Assessment of Urban Infrastructure in Hebei Province Based on Random Forest

Wang Yongxin[1], Dong Jiubo[1,*], Li Haijun[1]

[1]*Institute of Disaster Prevention Science and Technology, Sanhe, Hebei, China*
[*]*Corresponding author: dong15922164740@163.com*

Abstract: *Conducting a scientifically rigorous resilience assessment of urban infrastructure is crucial for maintaining the stability of essential urban functions. By analyzing the underlying meaning of urban infrastructure resilience, an evaluation index system was developed, encompassing five key dimensions: transportation, communication, water supply and drainage, sanitation, and energy. Leveraging index data from 2018 to 2022, training and testing datasets were generated using random sampling techniques. The classification model used for assessment was refined through feature selection and parameter optimization processes, and the enhanced model was subsequently applied to evaluate the resilience across the entire study region. The findings indicate the following: (1) The performance of the optimized model has improved across all five metrics, including the F1 score, with all values exceeding 0.94; (2) Upon assessing the resilience of each subsystem, the water supply and drainage resilience exhibited relatively stable trends within a narrow fluctuation range. Communication and energy resilience both demonstrated steady upward trajectories, whereas transportation and sanitation resilience followed a pattern of "decline – recovery – stabilization"; (3) Over the study period, the resilience of urban infrastructure in Hebei Province varied both temporally and spatially. The spatial distribution of the results revealed a pattern characterized by "higher values in central regions and lower values in northern and southern areas," accompanied by localized clustering effects. The overall regional resilience experienced fluctuations but showed improvement, particularly in northwest cities. Areas with fluctuating changes exhibited relatively larger discrepancies in resilience assessment outcomes. These methodologies and findings hold significant implications for the development of resilient cities and the strategic planning of urban infrastructure.*

Keywords: *Urban Infrastructure; Resilience; Random Forest Classification; Model Optimization; Hebei Province*

1. Introduction

Hebei Province, situated in northern China, is encircled by Beijing and Tianjin and faces the Bohai Sea to the east. Its strategic geographical position makes it a crucial component of the national initiative for coordinated development in the Beijing-Tianjin-Hebei region, contributing to the formation of a world-class urban cluster. Consequently, advancing new urbanization and promoting high-quality integrated construction are pivotal steps in implementing the "Outline Plan for Coordinated Development of the Beijing-Tianjin-Hebei Region" [1]. The province's geological and topographical characteristics are highly diverse and dynamic. Influenced by the typical summer monsoon climate, Hebei frequently experiences heavy rainfall, resulting in well-established river systems. Meanwhile, human activities related to production and daily life have intensified, leading to frequent occurrences of various disasters and accidents with significant impacts. Over the past two decades, the urbanization rate has surged dramatically, rising from 26.08% at the start of the century to 63.77%, reflecting rapid urban expansion. These disaster-related interference factors have created overlapping risks, exacerbating the challenges of risk prevention [2]. To ensure effective high-quality construction and coordination within the Beijing-Tianjin-Hebei urban agglomeration, evaluating and enhancing the resilience of urban infrastructure in Hebei Province holds substantial significance for decision-making in disaster prevention infrastructure planning and construction.

Urban infrastructure resilience assessment involves systematically evaluating a system's capacity to sustain essential functions, adapt to environmental changes, endure sudden disruptions, and facilitate functional restoration and enhancement within a disturbed context. As a critical component of urban

ISSN 2616-5872 Vol.7, Issue 3: 55-67, DOI: 10.25236/AJEE.2025.070308

resilience assessment, evaluating urban infrastructure resilience aids in identifying vulnerabilities within the facility network, ensuring the stable and reliable functioning of urban systems and their associated operations. Presently, the primary methodologies employed for analyzing urban infrastructure resilience encompass simulation approaches, empirical techniques, and index-based frameworks, among others. The index-based framework identifies influential indicators impacting urban infrastructure, establishes a risk evaluation system, refines it through weight allocation, and ultimately determines the risk level. Notably, scholars such as She Shuo [3] and Yan Weidong [4] have developed evaluation frameworks from dimensions like energy, transportation, drainage, network infrastructure, and health sanitation assessments. Despite variations in the academic community's definition of urban infrastructure boundaries [5], energy, water supply and drainage, and transportation systems are widely acknowledged as core components. Common techniques for determining weights within the index framework include the Delphi method [6], Analytic Hierarchy Process (AHP) [7], Entropy Method [8]. Frequently utilized assessment models consist of PSR [9], DSR [10], and DPSIR [11]. The index evaluation approach employs multi-dimensional metrics and influencing factors to characterize infrastructure resilience, aiding in pinpointing strengths and weaknesses. However, this method often oversimplifies the intricate interactions within the infrastructure system by treating them as a linear aggregation of indicators, thereby neglecting the system's inherent complexity and failing to comprehensively account for all relevant factors. Machine learning, leveraging nonlinear modeling, dynamic learning, and implicit relationship extraction, overcomes the linear simplification constraints of the index-based approach[12]. Random Forest, with its strengths in nonlinear modeling, high-dimensional data processing, noise tolerance, interpretability, and practical efficiency, is especially well-suited for addressing the challenge of quantifying resilience in complex urban-scale systems [13].

Drawing on the theory of urban infrastructure resilience, this study develops an evaluation framework that incorporates various dimensions, including transportation, communication, water supply and drainage, sanitation, and energy. Leveraging data from 2018 to 2022, a resilience assessment model for infrastructure is established, utilizing the random forest classification algorithm for its development, training, and testing phases. The model undergoes refinement by integrating the selected F1 score and recall rate metrics. Subsequently, it is employed to evaluate the resilience of urban infrastructure in Hebei Province, aiming to offer insights and recommendations for the effective planning of infrastructure and resource optimization in the region.

2. Research Theory and Methodology

2.1 Definition of Urban Infrastructure Resilience

The concept of "resilience," originating as a basic physical principle, was initially presented by Holling to characterize an ecosystem's capacity to preserve or recover its initial functions following a disturbance [14]. It was later incorporated into the 2013 New York City Resilience Master Plan. Urban infrastructure, which serves as the tangible, organizational, systematic, and service-based support for the structured progression of urban society and economy, significantly influences the organization of urban production and daily activities, as well as ensuring the dependable provision of essential public services [15].

Thus, urban infrastructure resilience can be understood as a city's capacity to efficiently withstand, adjust to, rebound from, and gain insights from the stresses and shocks brought by internal and external disasters, accidents, and other critical events. This ensures the continuity of essential urban functions and services while supporting the sustainable development of the socio-economy [16]. Consequently, evaluating urban infrastructure resilience can aid in pinpointing vulnerabilities and major risks within urban infrastructure systems. This holds substantial importance for refining infrastructure planning and construction decisions and strengthening the city's overall ability to withstand risks.

2.2 Methods for Optimizing and Assessing Random Forest Models

2.2.1 Basic Principles and Construction Process of Random Forest Algorithm

Random Forest is an ensemble learning method designed to improve the accuracy and stability of classification and regression tasks. By aggregating multiple decision trees (base classifiers), it addresses the limitations of a single decision tree. During its construction, the algorithm generates multiple training subsets through random sampling from the original dataset. Decision trees are then built on each subset, with features being randomly chosen at each split [17]. Finally, the predictions from

ISSN 2616-5872 Vol.7, Issue 3: 55-67, DOI: 10.25236/AJEE.2025.070308

all individual decision trees are combined to produce the overall output. The process of constructing decision trees relies on information gain to determine the best attribute for splitting, and the data continues to be divided until a stopping criterion is reached. For a dataset D containing n classes of samples, the proportion of its kth sample is denoted as Pk, and its information entropy is typically defined as:

$$\mathrm{Ent(D)} = -\sum_{k=1}^{n} p_k \, \mathrm{lb} P_k \tag{1}$$

In the formula, the smaller Ent(D) is, the more concentrated the category distribution in the dataset is; conversely, the more chaotic it is.

If the attribute set D is split into V subsets D_1, D_2, ..., D_V according to attribute a, the information entropies of these subsets will be Ent(D_1), Ent(D_2), ..., Ent(D_V). Consequently, upon incorporating attribute a to divide the dataset, the resulting entropy becomes:

$$\mathrm{Ent(D,a)} = \sum_{v=1}^{v} \frac{|D^v|}{|D|} \mathrm{Ent}(D^v) \tag{2}$$

Information gain indicates the difference in information entropy prior to and following the introduction of an attribute. The larger the gain, the more effective the outcome. It can be described as:

$$\mathrm{Gain(D,\ a)} = \mathrm{Ent(D)} - \sum_{v=1}^{v} \frac{|D^v|}{|D|} \mathrm{Ent}(D^v) \tag{3}$$

2.2.2 Random Forest Feature Selection

In the application phase, to improve model performance and simplify model complexity, this study aims to determine the optimal number of features and assess the importance of each feature using recursive feature elimination. Through evaluating correlations, an optimized feature set will be selected, which is expected to enhance training efficiency and improve the model's fitting performance.

(1) Recursive Feature Elimination

Recursive Feature Elimination determines the worst feature to be removed from the running results by calculating the average purity reduction of each feature in all decision tree node splits, and repeats the process of removing the remaining features and ranking their importance [18].

(2) Pearson Correlation Analysis

The Pearson correlation coefficient is primarily used to evaluate the linear association between two variables. When two variables share a linear relationship, their patterns of change will follow a specific consistency that can be expressed numerically through the correlation coefficient. The formula for this numerical expression is:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{4}$$

In the formula, and represent the observed values of two variables, while and denote their mean values. A value of r > 0 suggests a positive correlation, r < 0 implies a negative correlation, and r = 0 means there is no linear association. The closer the absolute value of R is to 1, the stronger the correlation; conversely, the closer the absolute value of r is to 0, the weaker the correlation.

2.2.3 Random Forest Hyperparameter Tuning

Adjusting the hyperparameters of a random forest, such as the number of decision trees and the maximum depth, can affect the model's complexity, diversity, and fit, thereby influencing its generalization ability. Considering the characteristics of common hyperparameter optimization, to avoid getting stuck in local optima and improve computational efficiency [19], the study intends to first use random search to roughly explore the parameter space and identify promising regions, and then conduct more refined optimization within these regions using grid search.

2.2.4 Evaluation Method of Random Forest Model

To effectively analyze and evaluate the performance and characteristics of the classification model and determine its classification effect on different categories, the confusion matrix method will be adopted, combined with accuracy, recall rate, precision rate, F1 value and other indicators to assess the generalization ability of the model. The horizontal direction represents the true classification, and the vertical direction represents the classification result of the model. The description of each indicator is as follows. The basic form of the confusion matrix for a four-classification problem is shown in Table

ISSN 2616-5872 Vol.7, Issue 3: 55-67, DOI: 10.25236/AJEE.2025.070308

1.

Table 1 The Basic Form of the Four-Class Confusion Matrix

	Category 1	Category 2	Category 3	Category 4
Category 1	TP1	FN2	FN3	FN4
Category 2	FP1	TP2	FN3	FN4
Category 3	FP1	FP2	TP3	FN4
Category 4	FP1	FP2	FP3	TP4

In the table, TPi denotes the count of samples that truly belong to the i-th category and are correctly predicted as belonging to the i-th category (true positive). FPi represents the count of samples that do not actually belong to the i-th category but are incorrectly predicted as belonging to it (false positive). FNi refers to the count of samples that truly belong to the i-th category but are mistakenly predicted as not belonging to it (false negative). Among these, accuracy indicates the ratio of samples that the model predicts correctly to the overall number of samples. For a four-class classification problem, the formula for calculating it is:

$$\text{Accuracy}=\frac{\sum_{i=1}^{4}(\text{TP}_i)}{\sum_{i=1}^{4}(\text{TP}_i+\text{FP}_i+\text{FN}_i)} \tag{5}$$

Recall measures the proportion of actual positive samples that the model successfully identifies, while precision measures the proportion of samples predicted as positive by the model that are actually positive.

The calculation formulas for the precision and recall of the i-th class are:

$$\text{Precision\_i}=\frac{\text{TP}_i}{(\text{TP}_i+\text{FP}_i)} \tag{6}$$

$$\text{Recall\_i}=\frac{\text{TP}_i}{(\text{TP}_i+\text{FN}_i)} \tag{7}$$

The F1 value is a performance metric that comprehensively considers precision and recall. The F1 value for the i-th class is calculated using the following formula:

$$\text{F1\_i}=\frac{2*(\text{Precision\_i}*\text{Recall\_i})}{(\text{Precision\_i}+\text{Recall\_i})} \tag{8}$$

3. Overview of the Study Area and Data Sources

3.1 Overview of the Study Area

Hebei Province is situated in the central and eastern regions of northern China, encircling the municipalities of Beijing and Tianjin. To the east, it borders the western and northern coasts of the Bohai Bay and administers the Xiongan New Area along with 11 prefecture-level cities. As a key component of the Beijing-Tianjin-Hebei urban cluster, Hebei benefits from notable geographical advantages. The landscape features higher elevations in the northwest and lower areas in the southeast. The northern and western sections consist of mountains and hills, whereas the southeastern part forms part of the North China Plain. The region experiences a temperate continental monsoon climate, characterized by clearly defined seasons and significant annual fluctuations in precipitation. The primary rivers within the Hai River system flow from the north and west toward the eastern coastal regions, often resulting in frequent flooding[20]. Located in the northern section of the North China Seismic Zone, the area is influenced by multiple fault systems trending NE-NNE and NW-SW, leading to strong tectonic activity and a high frequency of seismic events[21]. Nevertheless, rapid urbanization has triggered a sharp rise in population, uneven distribution, and heightened interdependence and vulnerability within urban systems, presenting substantial challenges to sustainable urban development. Consequently, assessing the resilience of urban infrastructure plays a crucial role in enhancing cities' disaster resistance capabilities, optimizing planning and resource allocation, and fostering sustainable growth.

3.2 Data source

This collection and collation covers 23 indicators across five dimensions - transportation, communication, water supply and drainage, sanitation, and energy - for all cities in Hebei Province from 2018 to 2022. Among them, the data for water supply and drainage, sanitation, and energy are

ISSN 2616-5872 Vol.7, Issue 3: 55-67, DOI: 10.25236/AJEE.2025.070308

sourced from the publicly available data of the Ministry of Housing and Urban-Rural Development; the transportation data are from the "China Urban Statistical Yearbook"; and the communication data are derived from the "China Construction Statistical Yearbook".

4. Construction of the Evaluation Index System

Urban infrastructure is generally categorized into two types: engineering infrastructure and social infrastructure. As defined in the "Standard for Basic Terminology of Urban Planning" (GB/T50280-98) [22], engineering infrastructure consists of six primary systems: transportation, water supply and drainage, communication, energy, disaster prevention, and urban environment. Among these, energy, water supply and drainage, and transportation are considered the core components. Social infrastructure,on the other hand, covers areas such as education, research, and healthcare. Despite growing recognition of the importance of social infrastructure, existing studies and planning practices predominantly emphasize engineering infrastructure [23-24]. Consequently, this study aims to integrate both engineering and social infrastructure. Drawing on relevant literature [25-26], it proposes a framework for evaluating the resilience of urban infrastructure across five dimensions: water supply and drainage, transportation, energy, environmental sanitation, and communication. The backbone of a city's efficient functioning and sustainable development is rooted in the integration of its critical infrastructure systems. The water supply and drainage system plays a vital role in ensuring safe water access for both residents and industries, managing sewage treatment, handling rainwater, and preserving environmental cleanliness. Meanwhile, the transportation network facilitates the movement of people and goods, maintaining social order and fostering regional economic and commercial growth. The energy system, including electricity and gas, supplies stable power to support urban infrastructure and daily life while driving industrial and economic progress. The sanitation system contributes to maintaining the city's appearance and public health through cleaning services and waste management, protecting residents' well-being and enhancing the city's reputation. Additionally, the communication system establishes a platform for information sharing, encouraging social cooperation and knowledge dissemination. These interconnected systems collectively form the foundation for ensuring a city operates safely, efficiently, and as an attractive place to live.

Based on a comprehensive analysis of domestic and international research progress, and following the principles of systematicness, scientificity, typicality and data availability, this study selected 24 positive indicators (see Figure 1 for details):

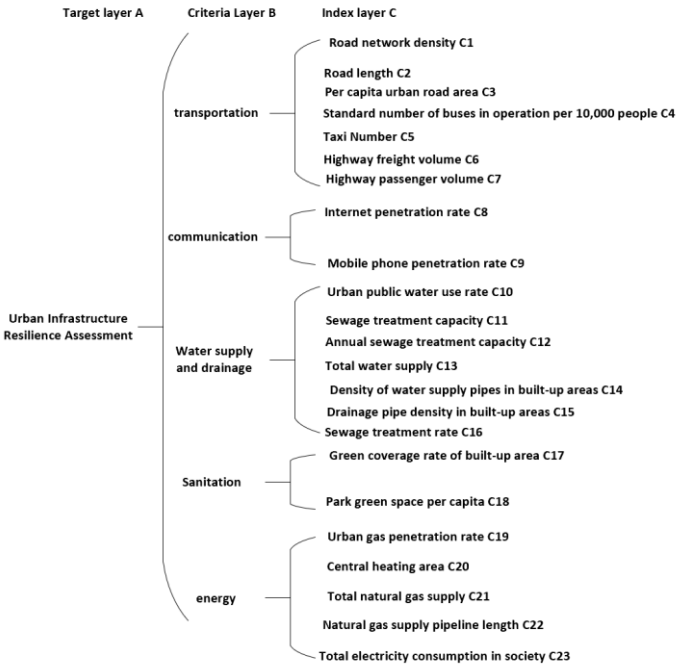

Figure 1: Assessment Framework for Urban Infrastructure Resilience

ISSN 2616-5872 Vol.7, Issue 3: 55-67, DOI: 10.25236/AJEE.2025.070308

5. Construction and Assessment of Random Forest Model

5.1 Data Processing

(1) Grading Standards for Urban Infrastructure Resilience

In order to create training and testing samples for each grade in a reasonable manner, this study applied the natural break method. This method categorized the data values of each urban infrastructure resilience indicator across cities in Hebei Province in 2018 into four distinct levels. The grading criteria for evaluating urban infrastructure resilience indicators are illustrated in Table 2.

Table 2 Assessment Criteria for Urban Infrastructure Resilience Indicators

Grade	Very High	High	Moderate	Low
C1/(Km.km$^{-2)}$	[98.30,100.00]	[96.77,98.29]	[92.73,96.76]	92.72
C2/Km	[47.01,148.60]	[26.31,47.00]	[22.01,26.30]	[18.50,22.00]
C3/Km	[24234.01,42792.00]	[14023.01,24234.00]	[6396.01,14023.00]	[4020.00,6396.00]
C4/(Vehicle/10^4 Person)	[15944.96,34274.79]	[6466.02,15944.95]	[4925.26,6466.01]	[4693.48,4925.25]
C5/Vehicle	[9.83,10.81]	[8.29,9.82]	[5.89,8.28]	[4.58,5.88]
C6/10^{4}T	[10.46,11.69]	[8.39,10.45]	[4.93,8.38]	4.92
C7/10^4 Person	[99.0,99.91]	[97.87,99.00]	[92.09,97.86]	92.08
C8/%	[8.09,8.55]	[7.36,8.08]	[6.48,7.35]	[5.83,6.47]
C9/%	[2012.87,2529.22]	[1142.98,2012.86]	[642.37,1142.97]	[525.31,642.36]
C10/%	[20.49,27.95]	[18.55,20.48]	[18.14,18.54]	[18.02,18.13]
C11/(10^{4}m^3)	[3.06,4.62]	[2.30,3.05]	[1.54,2.29]	[1.20,1.53]
C12/10^{4}m^3	[5207.01,8677.00]	[4204.01,5207.00]	[2863.01,4204.00]	[1470.00,2863.00]
C13/10^{4}m^3	[1470.00,2863.00]	[13770.01,24952]	[6697.01,13770]	[3870.00,6697]
C14/(Km.km^{-2})	[4022.01,9907.00]	[1842.01,4022.00]	[942.01,1842.00]	[768.00,942.00]
C15/(Km.km^{-2})	[37.37,42.86]	[29.70,37.36]	[24.48,29.69]	[21.87,24.47]
C16/%	[129.00,135.22]	[102.63,128.99]	[83.41,102.62]	83.4
C17/%	[44.01,47.26]	[42.32,44.00]	[40.38,42.31]	[38.88,40.37]
C18/m^2	[20.23,26.60]	[15.78,20.22]	[12.48,15.77]	[9.40,12.47]
C19/%	[99.73,100.00]	[98.86,99.72]	[96.20,98.85]	96.19
C20/10^{4}m^2	[8850.91,19765.10]	[5208.51,8850.90]	[3545.01,5208.50]	[2224.60,3545.00]
C21/10^{4}m^3	[78903.59,110890.82]	[48587.22,78903.58]	[13382.08,48587.21]	[4279.48,13382.07]
C22/km	[3100.01,5818.83]	[1982.01,3100.00]	[1058.89,1982.00]	[253.93,1058.88]
C23/10^{4}kWh	[5469998.01,8615362]	[2926301.01,5469998]	[1888415.01,2926301.00]	[1504480.00,1888415.00]

(2) Creation of Training and Testing Samples

In accordance with the references[27-28], based on Table 2, 5,000 data sets were randomly generated within each evaluation grade threshold. Subsequently, random interpolation was applied to produce a total of 20,000 data sets. These were then split into two subsets: 15,000 for training purposes and 5,000 designated as testing samples. The design of the urban infrastructure resilience evaluation samples and their anticipated outcomes is presented in Table 3.

Table 3 Sample Design for Urban Infrastructure Resilience Evaluation and Expected Output Design

Sample group	Target output	Evaluation level
1-5000	1	Level 1
5001-10000	2	Level 2
10001-15000	3	Level 3
15001-20000	4	Level 4

(3) Data Normalization Processing

In order to maintain the diversity of applicable models and remove the impact of varying measurement units within the data for subsequent machine learning modeling, this study applies extreme value normalization as a preprocessing method to the raw data. This approach effectively removes the effects of measurement units and ensures that all influencing factors are brought to a unified scale.

$$X^* = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{9}$$

In the equation, represents the normalized data, while X stands for the original data. To ensure that each evaluation index carries equal weight, is set to twice the upper limit threshold of each evaluation index, and is determined as one-tenth of the lower limit threshold of each evaluation index.

5.2 Assessment of Key Influencing Factors in the Model

5.2.1 Identification of Key Factors

The horizontal axis denotes the number of feature variables, while the vertical axis indicates the

ISSN 2616-5872 Vol.7, Issue 3: 55-67, DOI: 10.25236/AJEE.2025.070308

variation in model accuracy, showcasing the calculation outcomes of the recursive feature elimination method (see Figure 2). It is evident that when 15 feature variables are selected, the model achieves its peak accuracy, yielding the most optimal performance.

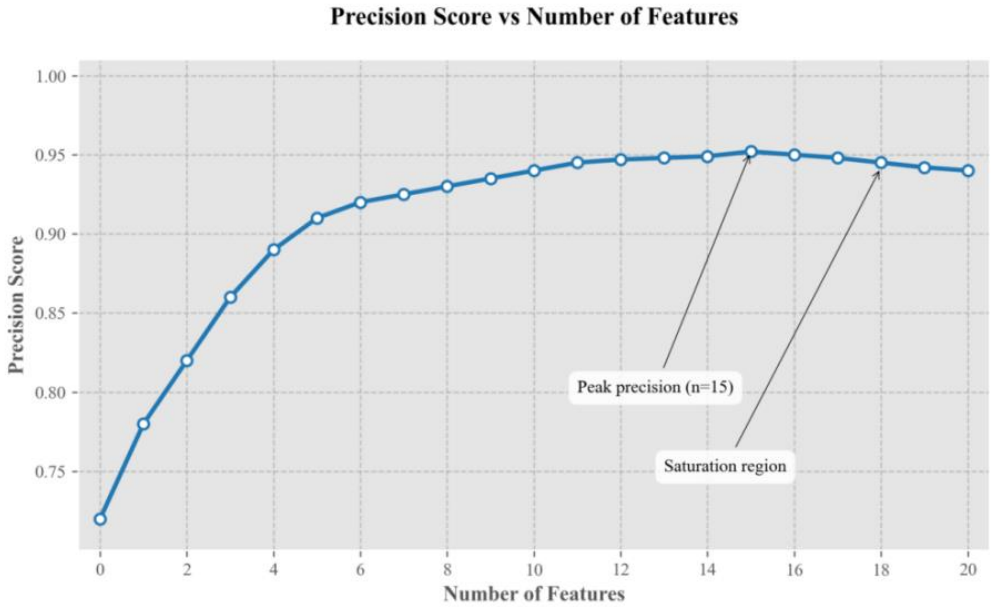

Figure 2: Relationship between the Number of Features and Model Accuracy

5.2.2 Key Factor Analysis

The significance of each feature was determined, and their importance rankings are presented as follows: C_{19} (0.15) holds the highest rank, followed by C_{16} (0.11). C_3 and C_9 (0.10) share the third position, while C_{10} (0.07) ranks lowest among the non-zero values. Subsequently, C_{15} and C_{22} (0.06), C_6 and C_{14} (0.05), C_2, C_8, and C_{23} (0.04), C4 and C_{17} (0.03), C_{13} and C_{18} (0.02), and finally C_1, C_5, C_{20}, and C_{21} (0.01) decrease progressively in importance. Notably, C_7, C_{11}, and C_{12} exhibit no importance, with a value of 0 (as shown in Figure 3).

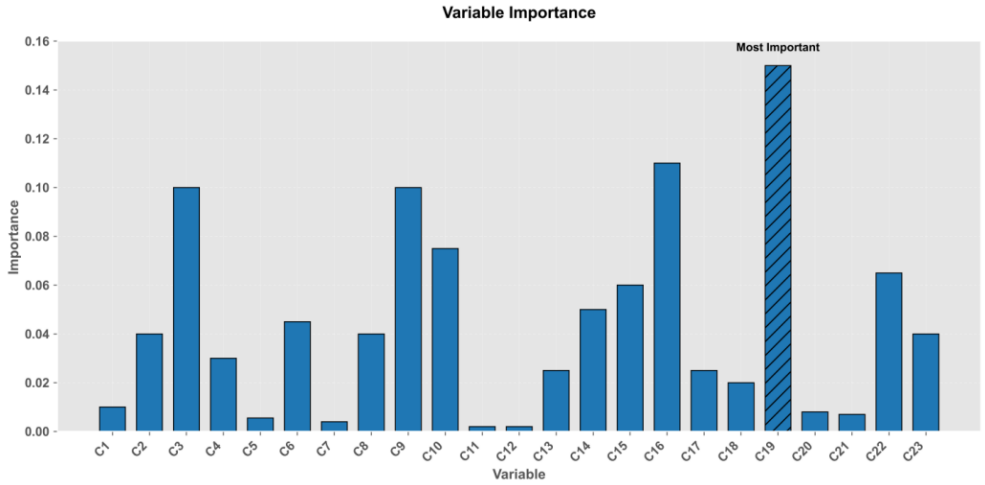

Figure 3: Values of the Importance of Each Feature

The correlation analysis heatmap presented in Figure 4 illustrates the relationships between different feature variables. In this visualization, darker shades represent stronger correlations, while lighter tones indicate weaker associations. The values span from -1 to 1, with 1 denoting a perfect positive relationship, -1 signifying a perfect negative relationship, and 0 reflecting no correlation at all. A higher absolute value closer to 1 suggests a stronger connection, whereas values nearer to 0 imply a weaker link. As depicted in Figure 4, C_3 and C_9, along with C_{16} and C_{19}, demonstrate a significant positive association. Meanwhile, C_5 and C_{12}, as well as C_7 and C_{11}, exhibit some level of negative correlation. The remaining features show relatively minimal correlations.

ISSN 2616-5872 Vol.7, Issue 3: 55-67, DOI: 10.25236/AJEE.2025.070308

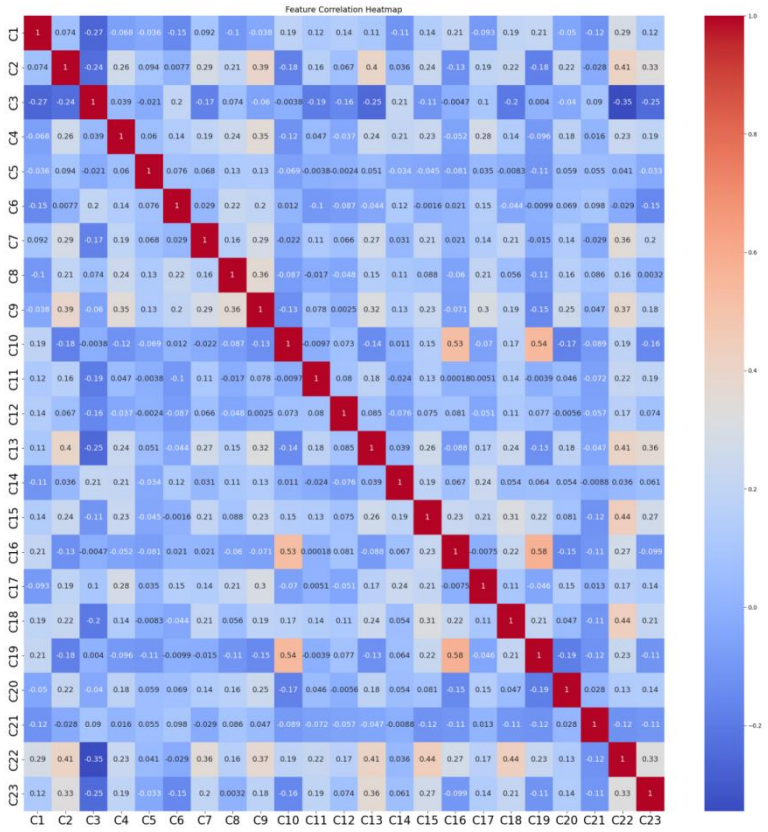

Figure 4: Heatmap of Correlation Analysis of Feature Variables

5.2.3 Ultimate Feature Determination

Following the determination of feature quantities, importance ranking, and correlation analysis, the recursive feature elimination method suggests that the optimal number of features is approximately 15. Subsequently, based on the analysis of feature importance, those with an importance value of 0 were removed. Additional correlation analysis was then performed for further screening. The final set of retained features comprises: C_1, C_2, C_3, C_4, C_6, C_8, C_9, C_{10}, C_{13}, C_{14}, C_{15}, C_{16}, C_{17}, C_{22}, and C_{23}. The excluded features include: C_5, C_7, C_{11}, C_{12}, C_{18}, C_{20}, and C_{21}.

5.3 Model for Optimizing Urban Infrastructure Resilience

5.3.1 Optimization of Model Parameters

The hyperparameters such as n_estimators, max_depth, min_samples_split, and min_samples_leaf directly affect the complexity, bias, variance, and final prediction performance of the model. First, set reasonable values for the above hyperparameters: n_estimators = [200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100], max_depth = [10, 20], min_samples_split = [2, 5, 10], min_samples_leaf = [1, 2, 4]. Use three-fold cross-validation to obtain the optimal values of each hyperparameter combination under random search: n_estimators: 400, max_depth: 10, min_samples_split: 2, min_samples_leaf: 1. Then, use grid search to arrange and combine the parameters generated by random search, and traverse these combinations. With the help of three-fold cross-validation, obtain the optimal values of each hyperparameter combination: n_estimators: 400, max_depth: 13, min_samples_split: 5, min_samples_leaf: 1.

5.3.2 Results of Model Assessment

The optimized inspection data model achieved an overall accuracy of 95.8%, along with a weighted F1 score of 94.9%. Specifically, the recall rates across the four categories were as follows: 97.5% for Category 1, 96.8% for Category 2, 97.3% for Category 3, and 97.1% for Category 4. Precision rates were recorded at 96.2%, 96.8%, 97.1%, and 97.3% respectively, while the corresponding F1 scores were 96.3%, 96.5%, 96.9%, and 97.0%. Additionally, the distribution of samples among the resilience levels was 24.88%, 25.50%, 25.08%, and 24.40%, respectively. The confusion matrix of the test data is shown in Table 4.

ISSN 2616-5872 Vol.7, Issue 3: 55-67, DOI: 10.25236/AJEE.2025.070308

Table 4 Confusion Matrix of Test Data

	1	2	3	4	Recall rate	Proportion
1	1120	0	0	0	97.5%	24.88%
2	5	1185	0	0	96.8%	25.50%
3	0	2	1168	0	97.3%	25.08%
4	0	0	0	1145	97.1%	24.40%
Accuracy rate	96.2%	96.8%	97.1%	97.3%	Accuracy rate: 95.8%	
F1 score	96.3%	96.5%	96.9%	97%	Weighted F1 score: 94.9%	

5.3.3 Effects of Model Optimization

Of these, prior to optimization (Baseline), the accuracy was 89.2%, precision was 88.9%, recall was 87.6%, and the F1 score was 88.3%. Following hyperparameter adjustment, each metric improved to 92.7%, 91.5%, 92.3%, and 91.9%, respectively. After combined optimization, the best performance was achieved, with an accuracy of 95.8%, precision of 94.7%, recall of 95.1%, and an F1 score of 94.9%. The model's performance saw varying degrees of improvement after feature selection and parameter fine-tuning. Figure 5 shows the final optimization result of the model.

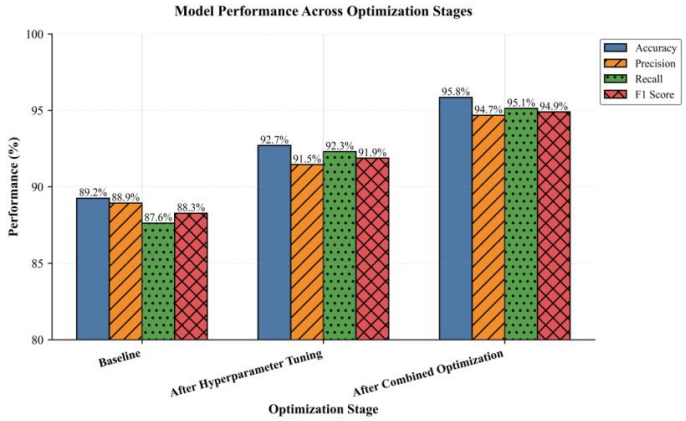

Figure 5: Optimization Outcome of the Model

6. Outcomes of Resilience Analysis

6.1 Time series analysis of single-system resilience

Figure 6 shows the time series analysis of five single systems，From the viewpoint of the water supply and drainage subsystem, several cities form a primary tier, including Shijiazhuang, Tangshan, Qinhuangdao, Handan, Xingtai, Zhangjiakou, and Cangzhou. These cities exhibit relatively stable fluctuations in their water supply and drainage resilience scores, with a steady trend in construction and management development. In contrast, Hengshui, Baoding, Langfang, and Chengde are positioned at a somewhat lower level. Notably, Hengshui has experienced a substantial rise in its water supply and drainage resilience score during the study period, increasing by 8.48%. The variations in water supply and drainage resilience among these prefecture-level cities remain within limited ranges, reflecting strong adaptability and stability.

From the viewpoint of the communication subsystem, Langfang City demonstrates remarkable performance, experiencing the highest increase in communication resilience from 2018 to 2022, with an 8.74% growth. Shijiazhuang City and Qinhuangdao City constitute a mid-level group, maintaining an average resilience value of roughly 0.14 during the five-year period. In comparison, the communication infrastructure resilience of other prefecture-level cities remains relatively underdeveloped. In general, the communication infrastructure resilience of all prefecture-level cities exhibits a stable upward trend, indicating ongoing investment and progressive enhancement in regional communication infrastructure development.

In terms of the transportation subsystem, Shijiazhuang and Tangshan lead the rankings, maintaining an average score above 0.35 over the past five years. In contrast, the other nine prefecture-level cities exhibit relatively weaker transportation resilience, with average scores remaining below 0.35. The overall development trend follows a "decline - recovery - stabilization" pattern. In 2020, the resilience

ISSN 2616-5872 Vol.7, Issue 3: 55-67, DOI: 10.25236/AJEE.2025.070308

scores for transportation infrastructure across all prefecture-level cities dropped by 6.002% compared to 2018, after which they started to recover gradually. Notably, Hengshui, Xingtai, and Qinhuangdao demonstrated more pronounced recovery progress.

From the viewpoint of the energy subsystem, Tangshan City, endowed with plentiful traditional energy resources like coal, oil, and natural gas, stands out with one of the highest energy infrastructure resilience scores among prefecture-level cities in Hebei Province. In contrast, the energy resilience levels of other prefecture-level cities in Hebei are relatively lower, averaging less than 0.16. Across all cities, there has been a consistent upward trend in energy resilience, reflecting ongoing enhancements in regional energy infrastructure development.

In terms of the sanitation subsystem, Shijiazhuang City ranks at the forefront, with an average resilience score exceeding 0.15 over the past five years. Zhangjiakou City, Chengde City, Qinhuangdao City, and Tangshan City constitute the second tier, sustaining an average score of approximately 0.135. The resilience of sanitation infrastructure in the remaining prefecture-level cities is relatively lower. From 2018 to 2022, the resilience of sanitation infrastructure in each prefecture-level city experienced an evolutionary process characterized by "decline - improvement - fluctuation - stability."

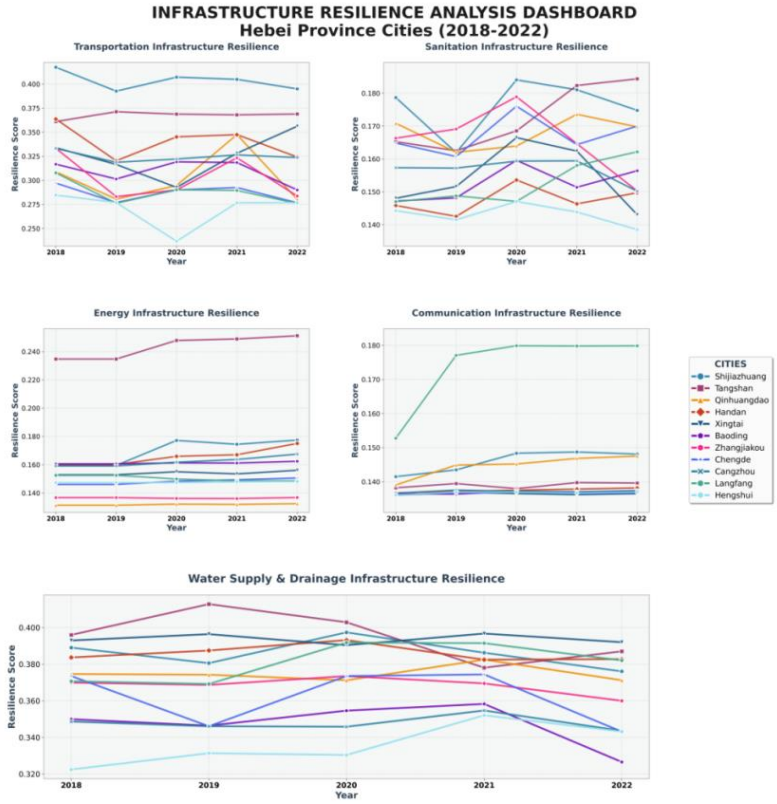

Figure 6: Time Series Chart of Urban Infrastructure Subsystem Resilience Levels

6.2 Characteristics of Spatial Differentiation in Compound Resilience

The composite infrastructure resilience scores for each city in Hebei Province were determined by calculating the weighted average of various indicators across the five subsystems: water supply and drainage, transportation, communication, environmental sanitation, and energy. For analysis, the results from 2018, 2020, and 2022 were chosen, and a spatio-temporal dynamic evolution map (Figure 7) was created. The natural break point method was applied to categorize the infrastructure resilience of Hebei Province's prefecture-level cities into four levels. The findings indicate that:

(1) During the early and middle stages of the study period, the overall spatial distribution of the assessment results generally exhibited a pattern of "high in the center and low in the north and south," accompanied by localized spatial clustering effects. In contrast, during the recent period, the spatial clustering of infrastructure resilience diminished, yet the originally high-resilience areas in the central region continued to maintain relatively high resilience levels.

ISSN 2616-5872 Vol.7, Issue 3: 55-67, DOI: 10.25236/AJEE.2025.070308

(2) Across the region, the overall trend of urban infrastructure resilience demonstrated a fluctuating upward trajectory. The proportion of cities with medium and high resilience levels increased over time, with the percentages of cities achieving medium and above resilience levels being 9.10%, 54.54%, and 81.81%, respectively. Notably, the resilience levels of cities such as Langfang, Zhangjiakou, and Shijiazhuang improved significantly compared to the earlier and middle stages. By 2022, Langfang and Zhangjiakou had advanced from low and medium resilience levels to high resilience levels, while Shijiazhuang progressed from a relatively high resilience level to a high resilience level.

(3) The disparities in resilience levels within the fluctuating change regions were pronounced. Cities like Cangzhou, Baoding, and Tangshan in the central area experienced fluctuations between medium and high resilience levels, whereas Xingtai's resilience oscillated between low and relatively low levels. Qinhuangdao displayed the most significant fluctuations, initially starting at a high resilience level before decreasing and subsequently rebounding. Meanwhile, the resilience levels of Hengshui and Chengde exhibited a slight downward trend.

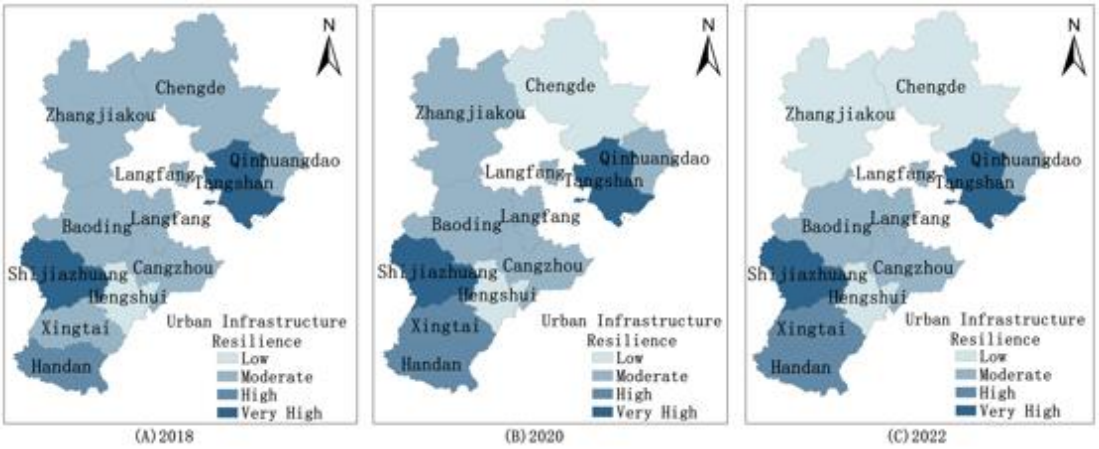

Figure 7: Spatial and Temporal Variations in Urban Infrastructure Resilience in Hebei Province

7. Conclusion

Evaluating the resilience of urban infrastructure plays a crucial role in recognizing the weaknesses within urban infrastructure systems and enhancing the development of resilient cities. Grounded in the theory of urban infrastructure resilience, this study establishes an index system comprising 23 indicators initially chosen from five key areas: transportation, energy, communication, sanitation, and water supply and drainage. Through the approach of randomly generating samples for both training and testing purposes, along with integrating hyperparameter tuning and feature selection to optimize the model, the refined model is utilized to assess the performance of urban infrastructure across all prefecture-level cities in Hebei Province from 2018 to 2022. The findings are summarized as follows:

(1) Following the process of feature selection and analysis, it was concluded that retaining 15 features yielded the highest model accuracy. By removing features with zero importance and high correlation, as well as fine-tuning the model parameters, there was a marked enhancement in model performance, with all evaluation metrics surpassing 0.94.

(2) The resilience performance across different subsystems exhibits variation. Water supply and drainage resilience remains generally stable, showing minimal fluctuations. Communication and energy resilience both demonstrate a consistent upward trend. In contrast, transportation and sanitation resilience initially decline, recover during the middle phase, and eventually stabilize in the later stages.

(3) The resilience of urban infrastructure in Hebei Province reveals pronounced spatiotemporal differentiation characteristics. Spatially, it forms a pattern characterized by higher values in the central region and lower values in the northern and southern areas, accompanied by localized clustering effects. Temporally, it displays an overall fluctuating upward trend, with particularly notable improvements observed in northwestern cities. The resilience assessment values differ significantly across regions with varying fluctuation patterns.

ISSN 2616-5872 Vol.7, Issue 3: 55-67, DOI: 10.25236/AJEE.2025.070308

References

[1] Han, Han, et al. "Spatiotemporal analysis of the coordination of economic development, resource utilization, and environmental quality in the Beijing-Tianjin-Hebei urban agglomeration." Ecological Indicators 127 (2021): 107724.
[2] Wang, Hefeng, et al. "Interaction between urbanization and eco-environment in Hebei Province, China." Sustainability 14.15 (2022): 9214.
[3] She Shuo, Duan Fang. Evaluation of Infrastructure Resilience Degree in the Yangtze River Delta Urban Agglomeration: Based on the TOPSIS Entropy Weight Method [J]. China Real Estate, 2020, (27): 31-35. DOI: 10.13562/j.china.real.estate.2020.27.006.
[4] Yan Weidong, Ding Chunlei, Wang Xiaolei. Research on the Evaluation of Urban Infrastructure Resilience Based on Entropy Weight-TOPSIS Model: A Case Study of Shenyang City [J]. Journal of Shenyang Jianzhu University (Social Science Edition), 2024, 26(01): 71-78.
[5] Sun Haiyan. Research on Evaluation and Enhancement Strategies of Urban Infrastructure Resilience [D]. Xi'an University of Architecture and Technology, 2022. DOI: 10.27393/d. cnki. gxazu. 2022. 000801.
[6] Linstone, Harold A., and Murray Turoff, eds. The delphi method. Vol. 1975. Reading, MA: Addison-Wesley, 1975.
[7] Orencio, Pedcris M., and Masahiko Fujii. "A localized disaster-resilience index to assess coastal communities based on an analytic hierarchy process (AHP)." International Journal of Disaster Risk Reduction 3 (2013): 62-75.
[8] Xiao, Wu, et al. "Ecological resilience assessment of an arid coal mining area using index of entropy and linear weighted analysis: A case study of Shendong Coalfield, China." Ecological Indicators 109 (2020): 105843.
[9] Fu, Xingfeng, et al. "A coupled PSR-based framework for holistic modeling and flood resilience assessment: A case study of the 2022 flood events in five southern provinces of China." Journal of Hydrology 636 (2024): 131255.
[10] Meng Xiaojing, Liu Chao, Cao Yingxue, et al. Research on the Spatio-Temporal Evolution of Urban Agglomeration Flood Disaster Resilience Based on Random Forest Optimization Algorithm [J]. Journal of Disaster Prevention and Mitigation Engineering, 2024, 44(04): 784-791. DOI: 10.13409/j.cnki.jdpme.20240131001.
[11] Zhao, Ruidong, et al. "Evaluating urban ecosystem resilience using the DPSIR framework and the ENA model: A case study of 35 cities in China." Sustainable Cities and Society 72 (2021): 102997.
[12] Derras, Boumédiène, and Nisrine Makhoul. "An overview of the infrastructure seismic resilience assessment using artificial intelligence and machine-learning algorithms." Proceedings of the ICONHIC (2022).
[13] Chen, Junfei, et al. "A machine learning ensemble approach based on random forest and radial basis function neural network for risk evaluation of regional flood disaster: a case study of the Yangtze River Delta, China." International journal of environmental research and public health 17.1 (2020): 49.
[14] Bo Jingshan, Wang Yuting, Bo Tao, et al. Research Progress on Resilient Cities and Suggestions for Resilient Urban and Rural Construction [J]. World Earthquake Engineering, 2022, 38(03): 90-100. DOI: 10.19994/j.cnki.WEE.2022.0060.
[15] Wang Ting, Xu Chuan. A Brief Discussion on Infrastructure Planning in Urban Planning [J]. Sichuan Architecture, 2010, 30(03): 15-16.
[16] Liu, Wei, and Zhaoyang Song. "Review of studies on the resilience of urban critical infrastructure networks." Reliability Engineering & System Safety 193 (2020): 106617.
[17] Belgiu, Mariana, and Lucian Drăguţ. "Random forest in remote sensing: A review of applications and future directions." ISPRS journal of photogrammetry and remote sensing 114 (2016): 24-31.
[18] Darst, Burcu F., Kristen C. Malecki, and Corinne D. Engelman. "Using recursive feature elimination in random forest to account for correlated variables in high dimensional data." BMC genetics 19 (2018): 1-6.
[19] Schratz, Patrick, et al. "Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data." Ecological Modelling 406 (2019): 109-120.
[20] Wang Guangpeng, Liu Lianyou, Hu Ziying. Research on Flood Disaster Risk Assessment at Grid Scale in the Beijing-Tianjin-Hebei Metropolitan Area [J]. Journal of Catastrophology, 2020, 35(03): 186-193.
[21] Xie Zhuojuan, Lü Yuejun, Fang Yi, et al. Research on Seismic Activity in the Beijing-Tianjin-Hebei Region [J]. Progress in Geophysics, 2019, 34(03): 961-968.
[22] Chen, Min, Chen, Jia, Guo, Xiangyu, Li, Yuxin, Zhu, Xiaowen, Zhang, Jianyun, et al. Measuring

ISSN 2616-5872 Vol.7, Issue 3: 55-67, DOI: 10.25236/AJEE.2025.070308

urban infrastructure resilience via pressure-state-response framework in four Chinese municipalities[J]. Applied Sciences, 2022, 12(6): 2819.
[23] Yin Hongling, Cui Dongxu, Wang Linshen, et al. Evaluation Objects, Perspectives and Index System of Regional Infrastructure [J]. Regional Research and Development, 2015, 34(06): 38-42.
[24] Xu W , Cong J , Proverbs D .Evaluation of infrastructure resilience[J].International Journal of Building Pathology and Adaptation, 2021.DOI:10.1108/ijbpa-09-2020-0075.
[25] Zhang Lequn, Chen Xuemin, Fu Xiaoyong, et al. Research on the Index System of Urban Infrastructure Construction in Lanzhou [J]. Urban and Rural Construction, 2012, (04): 54-55.
[26] Yang Hui. Research on the Integration of Infrastructure in 13 Cities of Beijing-Tianjin-Hebei Region Based on Coupling Coordination Degree Model [J]. Economics and Management, 2020, 34(02): 15-24.
[27] Cui Dongwen. Application of Support Vector Machine in the Identification of Lake and Reservoir Nutritional Status [J]. Water Resources Protection, 2013, 29(04): 26-30.
[28] Cui Dongwen, Jin Bo. Comprehensive Evaluation of Water Ecological Civilization Based on Random Forest Regression Algorithm [J]. Advances in Science and Technology of Water Resources, 2014, 34(05): 56-60 + 79.