Hindawi
International Journal of Geophysics
Volume 2023, Article ID 8532542, 15 pages
https://doi.org/10.1155/2023/8532542

Research Article

Evaluation of Building Seismic Capacity Based on Improved Naive Bayesian Algorithm

Yalong Li,[1,2] **Wei Wang**,[1] **Bin Tan**,[1] **and Hongxia Wang**[3]

[1]*Anhui Earthquake Agency, Hefei 230031, China*
[2]*Anhui Mengcheng National Geophysical Observatory, Mengcheng 233527, China*
[3]*Anhui Academy of Agricultural Sciences, Hefei 230031, China*

Correspondence should be addressed to Hongxia Wang; wanghongxia@aaas.org.cn

Received 6 July 2023; Revised 20 September 2023; Accepted 8 November 2023; Published 29 November 2023

Academic Editor: Arkoprovo Biswas

The influencing factors of building seismic capacity are analyzed, the basic cause events of the assessment target based on fault tree analysis (FTA) are determined, the basic cause events in the FTA model are classified and summarized, and a judgment system of building seismic capacity is built. The weight of each index factor in the Gini index calculation system is used, and the importance of the index is analyzed. On the basis of the Spearman correlation coefficient calculation of the index, the improved naive Bayesian algorithm is combined with the importance of the index to build a judgment model for the seismic capacity of housing buildings. The sample set is constructed based on the judgment system with the basic data of some housing buildings in Huoshan County. In order to improve the generalization ability and avoid overfitting, the K-SMOTE algorithm for mixed sampling was modified to improve sample balance, and random *k*-fold cross validation method was used for sample division and model optimization, achieving the determination of seismic capacity level of building. The research results indicate the following: (1) the accuracy of model evaluation is 93%, with model accuracy and recall rates of 0.913 and 0.93, respectively, indicating strong generalization ability of the model. (2) Selecting some actual examples of a building, the model judgment results are consistent with the actual results, verifying the correctness of the proposed method for building the model, which can be effectively used for determining the seismic capacity of building structures. (3) Applying the proposed method to the seismic capacity assessment of buildings in the Ta-pieh Mountains of Lu'an, it is concluded that the seismic capacity of urban buildings is common, while that of rural buildings is poor.

1. Introduction

In April 2022, China's "Fourteenth Five-Year" National Plan for Earthquake Preparedness and Disaster Reduction (hereinafter referred to as the "Plan") was released. The Plan proposes that by 2025, the level of earthquake disaster prevention will be significantly enhanced, the seismic capacity will be strengthened, the management of seismic fortification requirements will be strengthened according to law, and the strengthening of housing facilities in earthquake prone areas will continue. To achieve this goal, the first thing is to evaluate the seismic capacity of building in earthquake prone areas and, on this basis, implement targeted seismic reinforcement measures. The damage to building caused by earthquake is the main cause of casualties and economic losses. Using scientific method to evaluate the seismic capacity of building and determine the seismic capacity level of regional buildings has guiding significance for government department to strengthen the seismic fortification management of urban building. Domestic and foreign scholars have conducted relevant research on the evaluation of seismic capacity of building based on surveys and analysis. Document [1] discusses the construction and failure characteristics of five common types of buildings in Nepal and analyzes the seismic performance of different types of buildings. Document [2] determined the seismic capacity of buildings through nonlinear static analysis of three-dimensional numerical models. Document [3] uses a

trilinear analysis model to evaluate the bearing capacity of buildings. Document [4] evaluated the seismic capacity of some building structural models through nonlinear dynamic simulations. The research on the seismic capacity of building conducted above is mostly qualitative evaluation analysis, which effectively reflects the seismic capacity of building. However, the objective evaluation efficiency of the seismic capacity of regional building needs to be further considered. In order to objectively evaluate the seismic capacity of building structures, quantitative analysis of the properties of the building itself is necessary. Currently, there is no mature and complete quantitative indicator system in the seismic industry for evaluating the seismic capacity of building structures.

Domestic and foreign scholars have conducted research in different fields and achieved significant results in the analysis, establishment, and evaluation of indicator system. Documents [5–8] combine FTA method with AHP, FPN, ETA, and other methods to conduct research on railway safety risk, natural gas fire risk, and oil tank fire and explosion accident risk assessment. Documents [9–11] carried out safety assessment on chemical warehouse environment, safety of steel structural engineering, and risk of power grid tripping accident caused by typical natural disasters based on BN method. Documents [12–15] conduct risk assessment research in multiple fields such as supply chain, occupational safety, and business environment based on the AHP method. The above research uses quantitative deductive analysis to construct an indicator system, and on this basis, machine learning methods are used to construct an evaluation model, effectively conducting quantitative evaluation for specific problems. However, there are also certain limitations. The FTA method can trace the cause events based on the target events but cannot distinguish the overall safety situation of the target; AHP can analyze and evaluate the overall safety risk level of the target, but in the evaluation process, human subjective factors are heavy, and quantitative data is scarce, resulting in rough comparison, judgment, and calculation of results; BN is suitable for studying complex uncertainty problems, and the structural modeling process is complex.

In recent years, with the development of artificial intelligence, essential methods of AI collection, machine learning (ML), and computer vision (CV) have become increasingly popular in the field of building seismic evaluation. Document [16] proposes a machine learning-derived two-stage method for postearthquake building location and damage assessment considering the data characteristics of satellite remote sensing (SRS) optical images with dense distribution, small size, and imbalanced numbers; the multiscale features were successfully extracted and fused from SRS images of densely distributed buildings by optimizing the YOLOv4 model toward the network structures, raining hyperparameters, and anchor boxes. The fusion improved multichannel features, and the optimization of network structure and hyperparameters has significantly enhanced the average location accuracy of postearthquake buildings. Document [17] proposes a modified faster R-CNN for the multitype seismic damage identification and localization of RC columns; the RPN and fast R-CNN modules are merged into the proposed faster R-CNN by sharing their convolutional features to identify rectangular bounding boxes for multitype damage classification and localization. For the real-world damaged structural images containing complex background information, the proposed model autonomously drives attention to the damaged areas. Document [18] proposes a computer vision and machine learning-based seismic damage assessment framework for RC structures. A refined Park-Ang model is built to express the coupled effects of structural ductility and energy dissipation, which reflects the nonlinear seismic damage accumulation and generates a synthetical seismic damage indicator within 0~1 using hysteretic curve data.

The structural attributes of buildings are relatively complex, including not only quantitative indicators of specific values but also variable indicators described through semantic qualitative methods, and there is a certain correlation between different attributes. Naive Bayesian algorithm is an emerging classification method that can simultaneously process quantitative and qualitative information without being sensitive to missing data. It is widely used for predicting classification problems in various fields [19–22]. On the basis of analyzing the FTA method and taking into account the characteristics of building properties, this article uses the FTA method to clarify the target causal structure for determining the seismic capacity of building structures and constructs a building seismic capacity judgment system. The Spearman coefficient is used to calculate the correlation between indicators, and independent indicators are selected for AHP importance analysis and new attribute variables are screened. Design a method for evaluating the seismic capacity of building structures based on naive Bayesian algorithm. In order to improve the performance of the algorithm in handling imbalanced data, the K-SMOTE algorithm is used for mixed sampling to improve the balance of the sample set, and the random k-fold cross validation method is used to improve the naive Bayesian model, improving the accuracy and effectiveness of determining the seismic capacity of building. The determination method flow is shown in Figure 1. The method proposed in this article uses a quantitative approach to accurately determine the seismic capacity of building, which is more objective and can quickly determine the distribution of seismic capacity of regional buildings. It provides a theoretical basis for government departments to formulate regional disaster prevention and reduction strategies and implement seismic retrofitting of regional buildings and also provides reference for the preassessment of moderate to strong earthquake damage in the Ta-pieh Mountain area of Lu'an.

2. Construction of Evaluation System for Seismic Capacity of Building Structures

FTA is to select the target risk event as the top event and search for the direct cause and indirect cause events of the top event layer by layer from the top down to the basic cause events. The logical relationship between events is connected through logical symbols to form a directed fault tree analysis

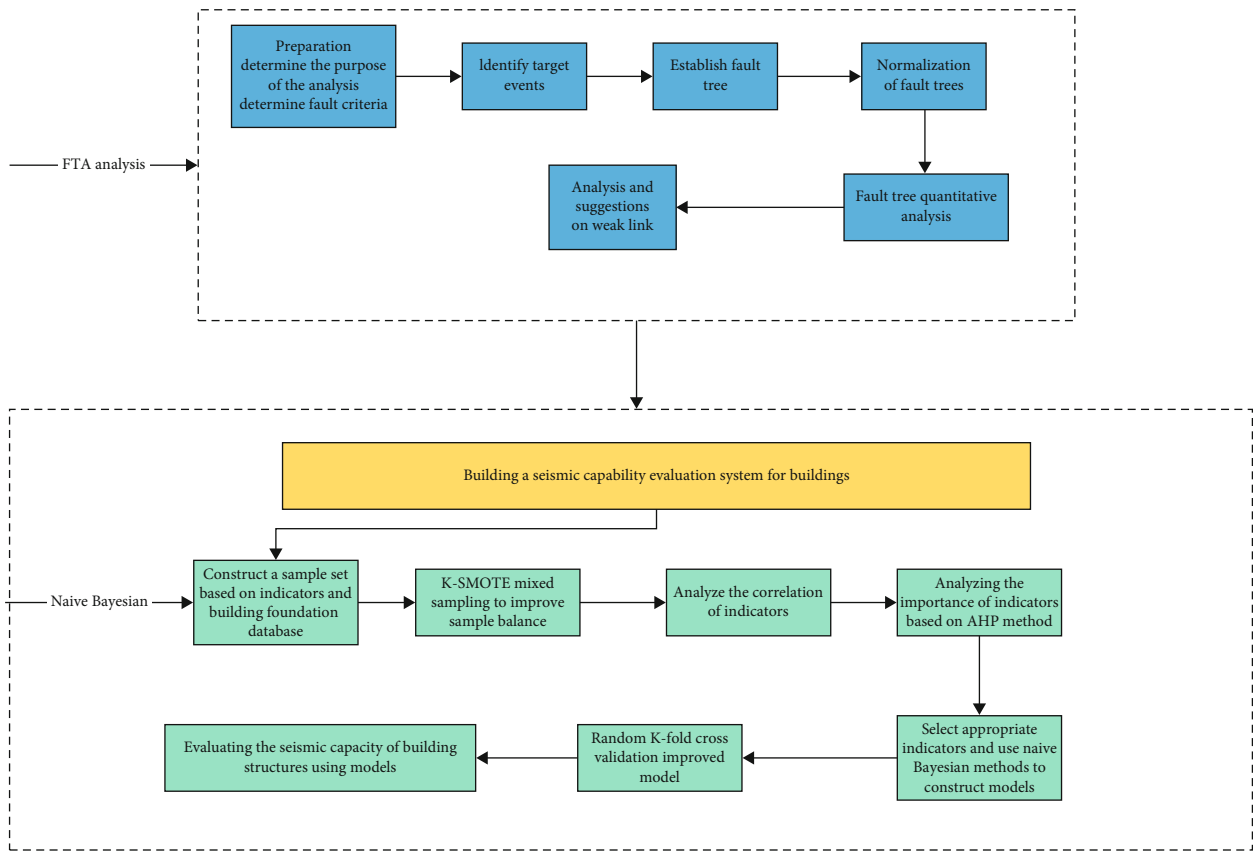

Figure 1: Evaluation process for seismic capacity of building.

model (FTA model) [23]. The FTA model can intuitively reflect the correlation between target risk events, intermediate events, and fundamental cause events, understand the causes of target risk occurrence, and find the best way to reduce risk, providing reference for developing risk control measures.

The evaluation of seismic capacity of building structures involves numerous kinds of information such as the structure of the building, environmental factors, and the degree of earthquake damage under different intensities. There are many influencing factors and a wide coverage. Factors at any level not only directly affect the seismic capacity level of the building but also have potential correlations with other factors, jointly determining the seismic capacity of the building. Therefore, in the process of analyzing the causes of top-level target events, comprehensive consideration should be given to the structural factors of the building, comprehensive environmental factors, personnel's awareness of earthquake prevention and disaster reduction, and the situation of building design specifications. The typical seismic damage characteristics of buildings produced by historical earthquakes can be used as the basis for the analysis of elementary event.

In fault tree analysis, the target event can be either an event that has already occurred or an expected event. The seismic capacity of building is taken as the target event. Based on historical earthquake damage data and structural characteristics of building, factors that affect the seismic capacity ability of the building are investigated, leading events and influencing factors are determined, and the analysis results show that the seismic capacity of the building is determined by the foundation, basic situation, comprehensive, residents' seismic awareness, standardization, and degree of earthquake damage. Each intermediate event is Logical-And relationship to the target event, and the root risk factors are gradually derived along the causal chain of the intermediate events to obtain the basic cause events. The intermediate event foundation is determined by the bearing capacity of the subgrade and the surrounding terrain, which are Logical-And relationship. The basic situation of intermediate events is determined by the foundation bearing capacity, seismic facilities, and column base connection, which are Logical-And relationship. The intermediate events are comprehensively determined by the bearing capacity of the upper structure, the number of floors, the age of the building, the integrated connection structure, the building structure, the setting of ring beams and structural columns, and the construction of the roof system, which are Logical-And relationship. The residents' seismic awareness in intermediate events is determined by the intensity of earthquake prevention and disaster reduction knowledge promotion and the presence of flammable and explosive materials in the building, which are Logical-And relationship. The standardization for intermediate events are determined by whether to refer to the design specifications for building foundation, whether to refer to the seismic design specifications for building, whether to implement the seismic

evaluation standard for building, and whether to implement the reliability evaluation standard for civil building, all of which are Logical-And relationship. The degree of earthquake damage caused by intermediate events is determined by the degree of earthquake damage in the VI degree area, the degree of earthquake damage in the VII degree area, and the degree of earthquake damage in the VIII degree area, which are Logical-And relationship. Based on the causal chain of events, determine the hierarchical relationship between the target event, intermediate event, and root node, and use logical gate to connect to form a fault tree, as shown in Figure 2. In addition, on-site sampling surveys were conducted to investigate the seismic capacity of some buildings in certain area, and the rationality of fault tree was checked from bottom to top.

On the basis of using FTA method to analyze the target events, follow the principle of systematicness and normalization, invite experts in the field of earthquake disaster assessment to evaluate the indicators of each floor, delete the indicators with low correlation, supplement the missing indicators, and finally build a set of building seismic capability judgment system consisting of 6 level 1 indicators and 21 level 2 indicators, as shown in Table 1.

3. Naive Bayesian Algorithm and Improvement

3.1. Naive Bayesian Algorithm. Naive Bayesian algorithm is a classification algorithm widely used in machine learning and data mining. It calculates the conditional probability of the sample to be judged to belong to each category based on Bayesian theorem and attribute conditional independence assumption and then judges it to the category with the largest probability.

Suppose there is a training sample set D, the number of samples is N, and there are m classes $C = \{C_1, C_2, \cdots, C_m\}$ in total. The data samples are represented by the n-dimensional feature vector $X = \{x_1, x_2, \cdots, x_n\}$, and $A_1, A_2, \cdots, A_n$ is the attributes of the samples. When there is a data sample X with an unknown label, the classification algorithm predicts that X belongs to the category with the highest posterior probability under X conditions. That is to say, the naive Bayesian classification algorithm assigns unknown samples to class C_i under the following conditions:

$$P(C_i|X) > P(C_j|X), \quad 1 \le i, j \le m, j \ne i. \tag{1}$$

According to the principle that each feature of naive Bayesian is independent from each other, the conditional probability of the sample X to be determined to belong to category C_i is calculated as

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}, \quad i = 1, 2, \cdots, m, \tag{2}$$

$$P(X|C_i) = \prod_{j=1}^{n} P(x_j|C_i), \quad j = 1, 2, \cdots, n, \tag{3}$$

$$P(C_i) = \frac{N_{C_i}}{N}. \tag{4}$$

Among them, x_j is the value of the pending sample X in attribute A_j, $P(X)$ is the joint distribution probability of $X = \{x_1, x_2, \cdots, x_n\}$, $P(x_j|C_i)$ is the conditional probability that the value of the sample X to be determined in attribute A_j is x_j on the premise that it belongs to category C_i, $P(C_i)$ is the prior probability of class C_i, and N_{C_i} is the number of samples of class C_i in the sample set.

If A_j is a discrete value attribute, then $P(x_j|C_i) = S_{ij}/N_{C_i}$, and S_{ij} is the number of samples in category C_i with a value of x_j under the condition of attribute A_j.

If A_j is a continuous value attribute, first assume that the attribute is subject to normal distribution, that is,

$$P(x_j|C_i) = \frac{1}{\sqrt{2\pi}\delta_{C_i}} \exp\left(-\frac{(x_j - u_{C_i})^2}{2\delta_{C_i}^2}\right). \tag{5}$$

Among them, u_{C_i} and $\delta_{C_i}^2$ are the mean and variance, respectively.

Classify the unknown sample X and calculate the probability that X belongs to each category as $P_i = P(C_i|X)$, $i = 1, 2, \cdots, m$, with $P = \{P_1, P_2, \cdots, P_m\}$. If the subscript of the maximum value in the set P is k, then X is classified as the k-th category with $k \in C$.

3.2. Algorithm Improvement. Naive Bayesian model has a solid theoretical foundation in statistics and relatively stable classification efficiency. The model requires very few estimated parameters and is not very sensitive to missing data, making the algorithm implementation relatively simple. The premise of naive Bayesian theory is to assume that attributes are independent of each other, but this assumption is often not valid in practical application, as there may be a large amount of redundancy between attributes, which can affect the classification efficiency of naive Bayesian model. When the correlation between attributes is small, naive Bayesian model exhibits better classification performance.

It should be pointed out that when using the naive Bayesian algorithm to process multidimensional imbalanced data, there are mainly the following shortcomings: (1) there are many influencing indicators for the seismic capacity of buildings, and different indicators contribute to different degrees of seismic capacity. There is mutual coupling between each indicator, creating data redundancy and increasing model complexity, reducing evaluation accuracy. (2) When dealing with imbalanced data, classification prediction tends to favor the majority class, which reduces the classification accuracy of the algorithm. (3) Naive Bayesian only uses the dataset once, and the final evaluation indicator calculated on the validation set is closely related to the original grouping. In some cases, the prediction performance may be poor due to the assumed prior model. Therefore, according to the above shortcomings of the current naive Bayesian model, we should optimize it from three aspects, the independence of attributes, the balance of sample data,

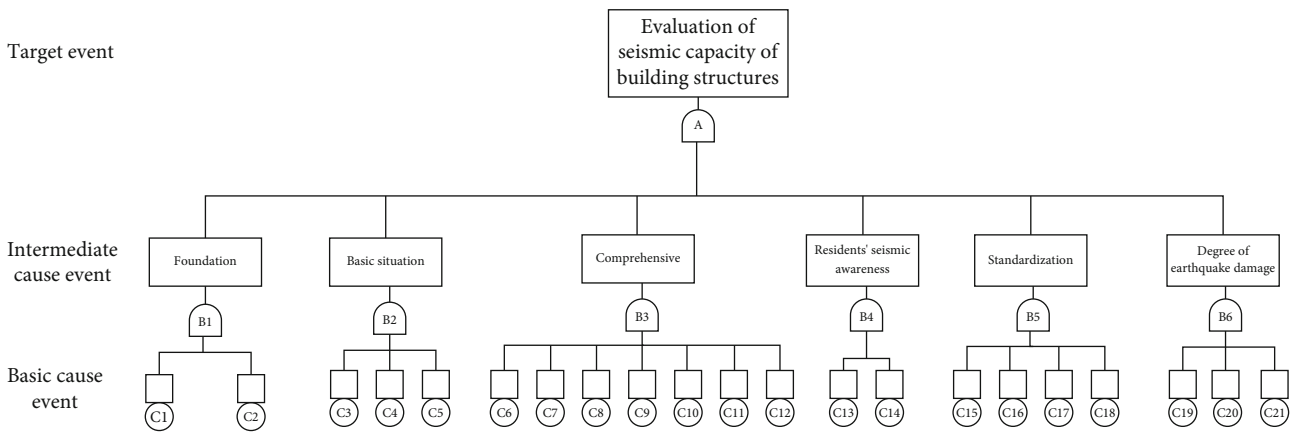

Figure 2: FTA model.

and avoiding overfitting of the model, and propose an improved naive Bayesian algorithm. The specific improvement ideas are as follows:

(1) Considering the impact of attribute correlation on the performance of naive Bayesian classification, appropriate improvements are made to the correlation between attributes. The AHP method is used to calculate the weights of each influencing indicator, and an appropriate number of indicators are selected to meet the independence assumption to a certain extent

(2) Preprocess the multidimensional imbalanced dataset using the K-SMOTE algorithm to achieve the goal of balancing the original dataset. Then, use random crossover validation to divide the balanced dataset into samples, construct new training and testing sets for model training, and improve the algorithm's generalization ability

(3) In the field of classification, although naive Bayesian algorithm can make relatively scientific judgments, some data must use subjective probabilities in the calculation process, which ultimately leads to deviation in the classification result. By training samples from multiple classifiers to enhance the learning ability of the algorithm, it is equivalent to handing the problem over to multiple classifiers to divide and conquer. Each classifier produces a result, and everyone votes together. When voting, the classification results of classifiers with low error rates should account for a larger proportion

3.2.1. AHP Calculation of Indicator Weights. AHP is a systematic and hierarchical analysis method that combines qualitative and quantitative analysis. It has practicality and effectiveness in dealing with complex decision-making problems [24]. Using the analytic hierarchy process to calculate the weights of each indicator in Figure 2, the approximate steps are as follows: (1) construct a judgment matrix, (2) consistency check of judgment matrix, and (3) calculate indicator weights.

The judgment matrix is the basis for conducting hierarchical analysis, which can be obtained by comparing the importance of the influence of criteria or indicator layer factors on the factors in the previous layer.

Let a certain criterion layer or indicator layer X have n factors, that is, $X = \{x_1, x_2, \cdots, x_n\}$. By comparing their importance in influencing the factors of the previous layer, the judgment matrix A is obtained as follows:

$$A = \left(a_{ij}\right)_{m\times n}. \tag{6}$$

In the formula, a_{ij} represents the comparison result of the i factor relative to the j factor.

In the process of comparing the importance of two factors, the importance of the basic cause event structure obtained from FTA and the experience of experts in various professional fields can serve as the basis for decision-making. The values of a_{ij} are shown in Table 2.

Calculate the maximum eigenvalues of each layer of the judgment matrix and the corresponding eigenvectors of each factor according to reference [25], and perform consistency testing on the judgment matrix. After passing the test, the eigenvectors (normalized) are used as the weight vectors for the influence of each factor in a certain layer on each factor in the previous layer. Finally, calculate the weights of each indicator based on the eigenvectors, and select appropriate indicators based on the indicator weights to form a new attribute dataset.

3.2.2. Hybrid Sampling Based on K-SMOTE Algorithm. The SMOTE algorithm [26] mainly involves randomly inserting artificial positive class samples into a few class samples to balance the distribution between classes and improve the classification performance of the classifier on the test set. The general execution process of the SMOTE algorithm is as follows: when processing positive class sample X, determine the k nearest neighbor samples closest to positive class sample X, and then, randomly select m samples from these k nearest neighbor samples. For each of these m samples X_i

Table 1: Evaluation system for seismic capacity of building.

Target	One-level indicator	Two-level indicator	Quantification of indicator
(A) Evaluation of seismic capacity of building	(B_1) Foundation	(C_1) Subgrade bearing capacity	High: 1/medium: 2/low: 3/poor: 4
		(C_2) Surrounding terrain	Favorable: 1/general: 2/disadvantageous: 3/dangerous: 4
	(B_2) Basic situation	(C_3) Foundation bearing capacity	High: 1/medium: 2/low: 3/poor: 4
		(C_4) Seismic facilities	Yes: 1/no: 2
		(C_5) Column base connection	Reasonable: 1/unreasonable: 2
		(C_6) Upper structure bearing capacity	High: 1/medium: 2/low: 3/poor: 4
	(B_3) Comprehensive	(C_7) Number of floors	Below 3 floors: 1/4~6 floors: 2/7~11 floors: 3/12 floors above: 4
		(C_8) Building age	Less than 30 years: 1/greater than or equal to 30 years: 2
		(C_9) Integrated connection structure	Reasonable: 1/unreasonable: 2
		(C_{10}) Building structure	Steel concrete: 1/brick concrete: 2/brick wood: 3/civil: 4
		(C_{11}) Setting of ring beams and structural columns	Reasonable: 1/unreasonable: 2
		(C_{12}) Roof system construction	Perfect: 1/general: 2/poor: 3
	(B_4) Residents' seismic awareness	(C_{13}) Intensity of publicity on earthquake prevention and disaster reduction knowledge	High: 1/general: 2/low: 3
		(C_{14}) Are there any flammable and explosive materials inside the building	Yes: 1/no: 2
	(B_5) Standardization	(C_{15}) Do you refer to the design specifications for building foundation	Yes: 1/no: 2
		(C_{16}) Do you refer to the seismic design specifications for buildings	Yes: 1/no: 2
		(C_{17}) Is the seismic evaluation standard for buildings implemented	Yes: 1/no: 2
		(C_{18}) Is the reliability evaluation standard for civil buildings implemented	Yes: 1/no: 2
	(B_6) Degree of earthquake damage	(C_{19}) Degree of seismic damage in the VI degree area	Collapse: 1/severe damage: 2/moderate damage: 3/minor damage: 4/basically intact: 5
		(C_{20}) Degree of seismic damage in the seventh degree area	Collapse: 1/severe damage: 2/moderate damage: 3/minor damage: 4/basically intact: 5
		(C_{21}) Degree of earthquake damage in the VIII degree area	Collapse: 1/severe damage: 2/moderate damage: 3/minor damage: 4/basically intact: 5

Table 2: Value and significance of a_{ij}.

a_{ij} value	Interpretation
1	The i factor is as important as the j factor
3	The i factor is slightly more important than the j factor
5	The i factor is significantly more important than the j factor
7	The i factor is more important than the j factor
9	The i factor is extremely important than the j factor
2, 4, 6, 8	The corresponding values of the intermediate states mentioned above
Reciprocal	The judgment value obtained by comparing the j factor with the i factor is $a_{ij} = 1/a_{ji}$

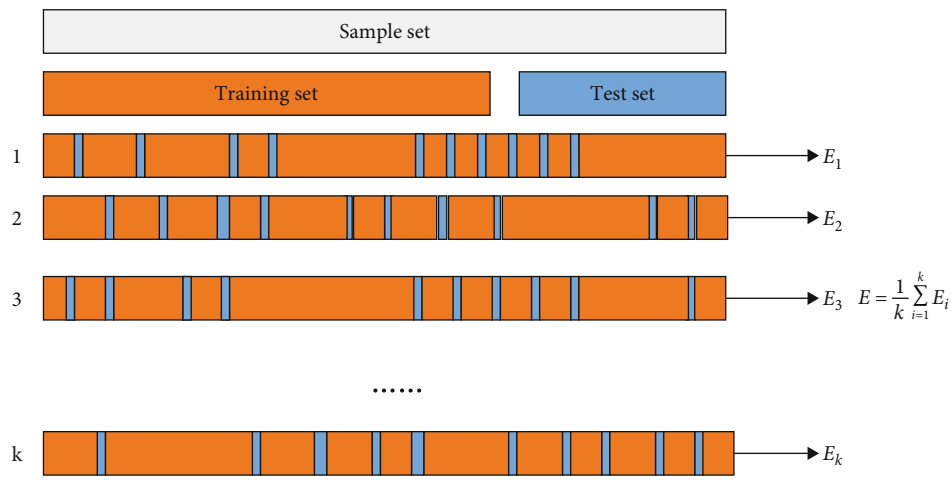

Figure 3: Random k-fold cross validation process.

Table 3: Standard for seismic capacity level of building.

Seismic capacity level	Seismic capacity	Description
1	Estimated seismic capacity meets the standard	When reaching VIII degree impact, the house is at a moderate degree of damage or below; when reaching level VII impact, the house is slightly damaged or below; after reaching VI degree impact, the house is basically intact
2	Suspected insufficient seismic capacity	When reaching level VII impact, the house is at a moderate level of damage or below; when reaching VI degree impact, the house is slightly damaged or below
3	Suspected severe lack of seismic capacity	When reaching VI degree impact, the house is at a moderate level of damage or below

$(i = 1, 2, \cdots, m)$, generate new artificial sample points according to the following:

$$X_{\text{new}} = X + \text{rand}\ (0, 1) * (X_i - X). \tag{7}$$

In the formula, X_{new} represents the newly interpolated sample, rand $(0, 1)$ represents the generation of a random number from 0 to 1, and the balanced dataset after mixed sampling using the SMOTE algorithm is D_{new}.

In response to the problem of boundary ambiguity in traditional SMOTE algorithm random sampling [27], in this paper, a clustering algorithm [28] was introduced to improve the SMOTE algorithm. The approximate implementation process is as follows: k-means is used to perform clustering operations on the entire sample space to determine the distribution status of each cluster in a minority class, and then, the clustering centers of each cluster are calculated. Interpolation is performed on the line between the cluster center and the sample within the cluster according to equation (8) to increase artificial sample data.

$$X_{\text{new}} = t_i + \text{rand}\ (0, 1) * (X - t_i), \quad i = 1, 2, \cdots, N. \tag{8}$$

Among them, $X \in t_i$ and X_{new} are the newly interpolated samples, t_i is the cluster center, and X is the original sample in the cluster with t_i as the cluster center.

3.2.3. Random k-Fold Cross Validation Improvement. The random k-fold cross validation improved the prediction performance of the model obtained only by training the known dataset to the unknown data, which is called "overfitting." The appearance of overfitting shows that the model did not learn the essential laws in the data. In the random forest algorithm, simple cross validation is used to avoid overfitting; for example, the samples are divided into two parts in proportion: 70% of the samples are used to train the model, and 30% of the samples are used for model validation, which to some extent improves the generalization ability of the algorithm. To address the shortcomings of simple cross validation, this paper proposes random fold cross validation to improve the algorithm. The steps are as follows:

(1) Firstly, the order of D_{new} in the balanced dataset is randomly disrupted, and then, the sample set is divided into k different combinations of training and testing sets

(2) The k sample sets are traversed in turn. Each time, the test set samples are used as the verification set, and all the other samples are used as the training

Table 4: Index weight of the evaluation system for seismic capacity of building.

	One-level indicator	Weight	Two-level indicator	Local weight	Global weight	Sort
Evaluation of seismic capacity of building	Foundation	0.0613	Subgrade bearing capacity	0.8	0.0490	5
			Surrounding terrain	0.2	0.0123	17
	Basic situation	0.1719	Foundation bearing capacity	0.1570	0.0270	12
			Seismic facilities	0.5936	0.1020	3
			Column base connection	0.2494	0.0429	6
	Comprehensive	0.6199	Upper structure bearing capacity	0.0971	0.0602	4
			Number of layers	0.0450	0.0279	11
			Building age	0.0293	0.0181	15
			Integrated connection structure	0.2212	0.1371	2
			Building structure	0.5330	0.3304	1
			Setting of ring beams and structural columns	0.0293	0.0181	14
			Roof system construction	0.0451	0.0279	10
	Residents' seismic awareness	0.0286	Intensity of publicity on earthquake prevention and disaster reduction knowledge	0.3333	0.0095	18
			Are there any flammable and explosive materials inside the building	0.6667	0.0190	13
	Standardization	0.1183	Do you refer to the design specifications for building foundation	0.2390	0.0283	9
			Do you refer to the seismic design specifications for buildings	0.3397	0.0403	7
			Is the seismic evaluation standard for buildings implemented	0.2808	0.0333	8
			Is the reliability evaluation standard for civil buildings implemented	0.1405	0.0167	16

set. The naive Bayesian algorithm is used to train and evaluate the model

(3) Finally, use the average of k evaluation indicators as the final evaluation indicator

By setting a random number seed to ensure the same number of samples for each data split, the consistency of cross validation is ensured. The sample set splitting and training process is shown in Figure 3.

This method uses a random approach to construct a sample set, with the model calculation results obtained from each sample set represented by E_i, and the average of k evaluation indicators was used as the final evaluation value after cross validation. Random k-fold cross validation can encourage the model to learn samples from multiple aspects and avoid falling into local extremum.

4. Evaluation of Seismic Capacity of Building

4.1. Source of Sample Data. The data source of this paper is the basic database of housing construction in Huoshan County, and a dataset is constructed based on the seismic capacity evaluation system of housing buildings in this paper. Organize a research team to select urban and rural areas in Huoshan to conduct a sampling survey on the properties of houses and buildings, and complete the missing data of indicator $C_1 \sim C_{12}$. Update indicator $C_{13} \sim C_{18}$ data by inquiring with the housing and construction department and street communities, filling out questionnaires, and other forms to form basic data on the seismic capacity status of houses in villages/towns/counties. Invite three experts with rich experience in earthquake damage to evaluate the seismic damage situation of sampled buildings under different intensities on site, and determine the seismic capacity of the buildings. Based on the results of averaging the three experts, the seismic capacity level of the buildings was obtained, and the indicator $C_{19} \sim C_{21}$ was supplemented. Finally, 326 data samples were formed. The seismic capacity level standards for building structures are shown in Table 3.

There are a total of 326 training and testing samples, each containing 22 attributes (secondary indicators and seismic fortification level), classified by seismic fortification level labels. The specific number of samples is as follows:

(1) Estimated seismic capacity meets the standard, including 29 cases (8.9%)

(2) Suspected insufficient seismic capacity, including 97 cases (29.7%)

(3) Suspected severe insufficient seismic capacity, including 200 cases (61.4%)

To ensure the balance of various samples in the sample set, the K-SMOTE algorithm is used to perform mixed

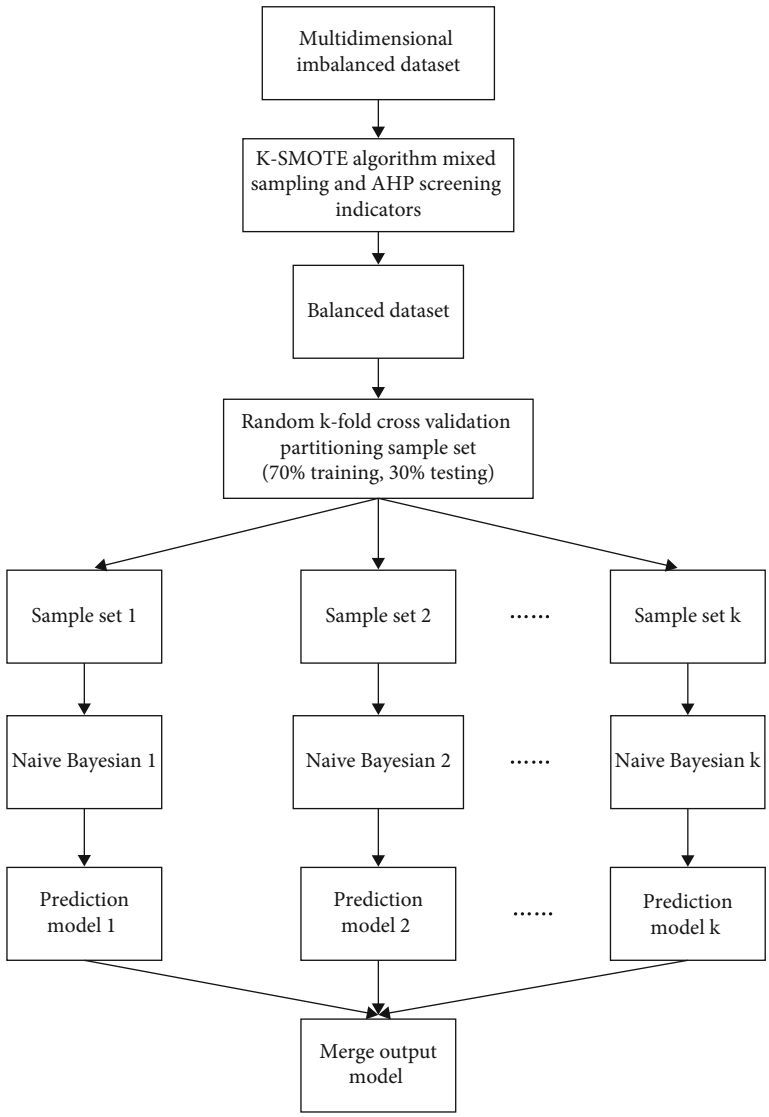

Figure 4: Improved naive Bayesian algorithm model construction process.

Table 5: Model evaluation form.

Seismic capacity level	Accuracy rate	Recall rate	***F*1** score
1	0.876	0.92	0.897
2	0.87	0.93	0.899
3	0.862	0.94	0.899

sampling on the sample set, and a new dataset with a sample size of 600 is constructed. Specifically, the estimated seismic capacity of 200 samples meets the standard. 200 samples are suspected to have insufficient seismic capacity, and 200 samples are suspected to have severely insufficient seismic capacity, achieving a balance of various sample data.

4.2. Indicator Correlation and Weight Analysis. Construct a set of indicators $C = \{C_1, C_2, \cdots, C_n\}$, $n = 21$, where C_i is the second-level indicator in the evaluation system for seismic capacity of building structures. In the naive Bayesian algorithm, it is necessary to maintain the independence of each indicator and select the secondary indicators in the seismic capacity evaluation system of buildings for the Spearman correlation coefficient analysis. The calculation formula is as follows:

$$r_s = 1 - \frac{6\sum_{j=1}^{n}\left(R_j - Q_j\right)^2}{n\left(n^2 - 1\right)}. \tag{9}$$

The paired values of two indicators x and y are ranked in order from small to large (or from large to small). $x, y \in C$, x_j, and y_j represent the j-th value of the indicator sample values after ranking, R_j represents the rank of x_j, Q_j represents the rank of y_j, Q_j is the difference between the ranks of x_j and y_j, and n is the sample size. Find the features with the highest correlation with the target value, and the correlation between these

Table 6: Actual sample prediction results.

Number	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}	C_{13}	C_{14}	C_{15}	C_{16}	C_{17}	C_{18}	C_{19} prediction	C_{20} prediction	C_{21} prediction	Prediction of seismic capacity	Expert actual judgment
1	3	1	4	1	2	3	2	1	1	1	1	3	3	2	2	1	1	1	5	5	4	1	1
2	3	3	3	2	2	2	1	1	1	2	2	3	3	2	2	1	1	2	4	4	3	2	2
3	4	2	3	2	1	4	3	1	2	3	2	3	2	2	2	2	2	2	3	2	1	3	3
4	2	2	3	2	2	3	1	1	1	4	2	2	3	2	2	2	2	2	2	2	1	3	3
5	4	3	3	2	1	2	2	1	2	4	2	3	3	2	2	2	2	2	2	1	1	3	3

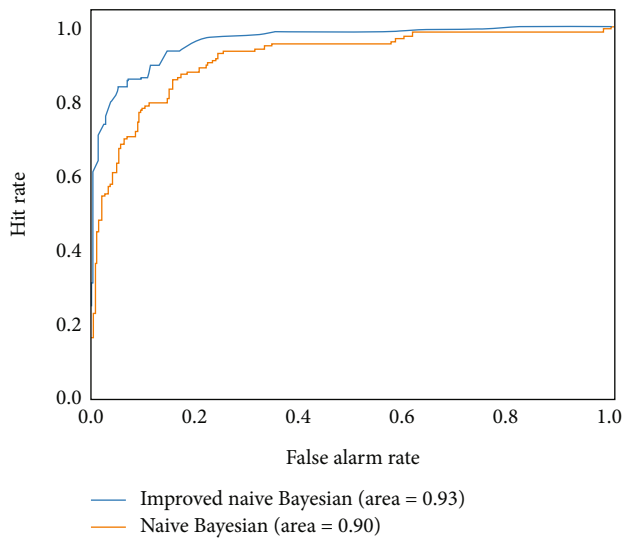

Figure 5: ROC curve comparison chart.

features is lower. The correlation matrix of each indicator feature is

$$R = \begin{bmatrix} 1.000 & 0.008 & -0.033 & 0.079 & -0.010 & \cdots & 0.103 & -0.040 & 0.002 & -0.014 & -0.005 \\ 0.008 & 1.000 & 0.204 & 0.158 & 0.004 & \cdots & -0.026 & -0.050 & -0.047 & -0.019 & 0.008 \\ 0.033 & 0.204 & 1.000 & 0.100 & -0.056 & \cdots & -0.070 & 0.048 & -0.036 & 0.005 & -0.052 \\ 0.079 & 0.158 & 0.100 & 1.000 & 0.107 & \cdots & 0.031 & 0.068 & -0.150 & -0.094 & -0.136 \\ -0.010 & 0.004 & -0.056 & 0.107 & 1.000 & \cdots & 0.038 & -0.066 & 0.007 & -0.018 & 0.019 \\ -0.084 & 0.086 & 0.089 & -0.163 & -0.103 & \cdots & -0.059 & -0.008 & 0.079 & -0.017 & 0.009 \\ -0.031 & -0.072 & 0.019 & -0.082 & 0.008 & \cdots & 0.016 & 0.038 & 0.051 & 0.018 & -0.080 \\ 0.030 & 0.015 & -0.069 & 0.022 & 0.064 & \cdots & -0.050 & 0.063 & -0.060 & -0.060 & -0.058 \\ 0.037 & -0.053 & -0.045 & -0.101 & 0.040 & \cdots & 0.049 & -0.062 & 0.030 & -0.006 & 0.045 \\ -0.044 & -0.032 & 0.033 & 0.045 & 0.008 & \cdots & 0.137 & 0.137 & -0.587 & -0.688 & -0.776 \\ -0.081 & -0.016 & -0.086 & -0.060 & 0.051 & \cdots & -0.070 & 0.117 & -0.057 & -0.045 & -0.011 \\ -0.014 & 0.012 & 0.102 & 0.023 & -0.097 & \cdots & 0.043 & -0.040 & -0.022 & -0.021 & 0.019 \\ -0.138 & 0.002 & -0.023 & -0.054 & -0.060 & \cdots & 0.040 & -0.057 & -0.005 & 0.013 & -0.030 \\ 0.006 & -0.094 & -0.091 & 0.025 & -0.062 & \cdots & -0.029 & 0.114 & -0.093 & -0.085 & 0.002 \\ -0.047 & 0.013 & -0.132 & 0.074 & 0.014 & \cdots & -0.017 & 0.061 & -0.138 & -0.036 & 0.012 \\ -0.043 & 0.037 & 0.066 & 0.003 & -0.076 & \cdots & 0.047 & 0.063 & -0.112 & -0.123 & -0.092 \\ 0.103 & -0.026 & -0.070 & 0.031 & 0.038 & \cdots & 1.000 & -0.032 & -0.224 & -0.204 & -0.168 \\ -0.040 & -0.050 & 0.048 & 0.068 & -0.066 & \cdots & -0.032 & 1.000 & -0.183 & -0.243 & -0.224 \\ 0.002 & -0.047 & -0.036 & -0.150 & 0.007 & \cdots & -0.224 & -0.183 & 1.000 & 0.732 & 0.580 \\ -0.014 & -0.019 & 0.005 & -0.094 & -0.018 & \cdots & -0.204 & -0.243 & 0.732 & 1.000 & 0.715 \\ -0.005 & 0.008 & -0.052 & -0.136 & 0.019 & \cdots & -0.168 & -0.224 & 0.580 & 0.715 & 1.000 \end{bmatrix}. \tag{10}$$

R_{mn} represents the correlation coefficient matrix between indicator C_m and indicator C_n. According to equation (10), it can be seen that the building structure (C_{10}) has a strong correlation with the seismic damage degree (C_{19}) in the VI degree zone, the seismic damage degree (C_{20}) in the VII degree zone, and the seismic damage degree (C_{21}) in the VIII degree zone. The correlation between the three indicators $C_{19} - C_{21}$ is high, and the indicator $C_{19} - C_{21}$ is a nonbasic attribute indicator of building construction. To ensure the independence of the attribute set, select the indicator $C_1 - C_{18}$ and use the AHP method to assign weights and sort them according to the global weights of the secondary indicators, as shown in Table 4.

From Table 4, it can be seen that in the determination of seismic capacity of building, the proportion of indicator building structures is the highest, followed by overall connection structures, seismic facilities, upper structure bearing capacity, and foundation bearing capacity. Overall, the seismic capacity of building structures is greatly influenced by three levels: foundation, comprehensive, and construction standards. The weight of the two indicators, namely, the promotion intensity of earthquake prevention and disaster reduction knowledge and the surrounding terrain, is relatively small. In the management of seismic fortification, emphasis should be placed on indicators with high weights, and preventive measures should be taken to prevent risks from escalating. After analyzing the correlation of indicators and ranking their weights, the top 16 indicators were finally extracted as a new attribute set, and then, a naive Bayesian model was constructed for the new attribute dataset.

4.3. Model Building. Process the data from the sample data source, construct a sample set according to evaluation indicators, obtain a balanced dataset based on K-SMOTE algorithm mixed sampling, use AHP method to filter indicators to form a new attribute dataset, and then, use random k-fold cross validation to generate k different sample sets. In the experiment, $k = 5$, use naive Bayesian algorithm to construct independent models for each sample set, use the constructed model to classify and predict the test sets in each sample set, and average the results of each model to obtain the seismic damage degree and seismic capacity level of the building under different intensities. The model construction process is shown in Figure 4.

4.4. Experimental Results

4.4.1. Model Evaluation. Evaluate the model using classification accuracy, precision, and recall, and the calculation results are shown in Table 5.

The accuracy rate in Table 2 indicates that the evaluation result of the model is the proportion of samples at that level, which is actually the proportion of samples at that level. The average accuracy rate of this model evaluation is 0.87, indicating the effectiveness of the model for sample classification. The recall rate represents the actual proportion of samples at that level, and the evaluation result is the proportion of samples at that level. For seismic fortification capability levels 1, 2, and 3, the recall rate is above 90%, with an average recall rate of 0.93, reflecting the model's ability to find relevant samples in a given sample set; $F1$ score is an indicator used in statistics to measure the accuracy of a model. It takes into account both the accuracy and recall of the classification model. $F1$ scores for seismic fortification capability levels 1, 2, and 3 are all above 0.89, reflecting the strong evaluation ability of the model for samples. The overall accuracy of the model evaluation is 93%, which verifies the effectiveness of the method proposed in this article in determining the seismic capacity of building structures.

Table 7: Calculation results of seismic capacity of some building areas in Ta-pieh Mountains of Lu'an.

Name	Proportion of buildings with standard fortification standards	Suspected insufficient construction proportion	Suspected severe shortage of buildings	Regional seismic performance index	Regional seismic capacity level
Shuanghe Town, Jinzhai County	0.0039	0.1302	0.8659	0.0756	Poor
Meishan Town, Jinzhai County	0.1211	0.3658	0.5131	0.0980	Common
Youfangdian Township, Jinzhai County	0.0006	0.0547	0.9447	0.0723	Poor
Taoling Township, Jinzhai County	0.0000	0.0059	0.9941	0.0702	Poor
Gubei Town, Jinzhai County	0.0002	0.1933	0.8064	0.0778	Poor
Zhufo'an Town, Huoshan County	0.1088	0.1839	0.7072	0.0893	Poor
Luoerling Town, Huoshan County	0.0470	0.6324	0.3206	0.1005	Common
Taipingfan Township, Huoshan County	0.0154	0.0000	0.9846	0.0717	Poor
Hengshan Town, Huoshan County	0.4900	0.2134	0.2966	0.1324	Common
Foziling Town, Huoshan County	0.0158	0.1893	0.7948	0.0793	Poor
Dongxixi Township, Huoshan County	0.0000	0.3546	0.6454	0.0842	Poor
Zhouji Town, Huoqiu County	0.0925	0.4866	0.4210	0.0996	Common
Madian Town, Huoqiu County	0.1484	0.2989	0.5528	0.0983	Common
Xihu Township, Huoqiu County	0.0025	0.0141	0.9834	0.0708	Poor
Xiadian Town, Huoqiu County	0.1551	0.1164	0.7285	0.0917	Common
Huhu Town, Huoqiu County	0.0653	0.1340	0.8007	0.0825	Poor

Table 6 shows the prediction of seismic capacity of some buildings.

4.4.2. ROC Analysis. For classification models, we are not only concerned with the accuracy of their predictions but also with the following two indicators: hit rate (the ratio of predicted x in all samples with a classification level of x) and false alarm rate (the ratio of predicted x in all samples with an actual classification level of x), which are used to evaluate the model through the ROC curve (receiver operating characteristic curve) drawn by both. We hope that under the same threshold, the false alarm rate should be as small as possible, and the hit rate should be as high as possible; that is, the ROC curve should be as steep as possible, and the corresponding AUC (area under ROC curve) value should be as high as possible. In order to further verify the evaluation performance of the proposed method, the improved naive Bayesian method and the traditional naive Bayesian method were compared and analyzed through ROC curves, as shown in Figure 5.

From the figure, it can be seen that the area under the curve of the improved naive Bayesian method and the traditional naive Bayesian method is both greater than 0.90, indicating that the naive Bayesian model is feasible in determining the seismic capacity of buildings. Compared with the traditional naive Bayesian method, the ROC curve of the improved naive Bayesian method is closer to the upper left corner, and the evaluation accuracy is higher.

4.4.3. Application of Seismic Capacity Assessment for Regional Housing Buildings. The Ta-pieh Mountains are located in the eastern region of China, extending eastward to the Huoshan in Lu'an. Affected by the Tanlu Fault Zone, the frequency of small- and medium-sized earthquakes in the Dabie Mountain region in Lu'an remains high. According to earthquake records in Anhui Province, there have been 8 moderate to

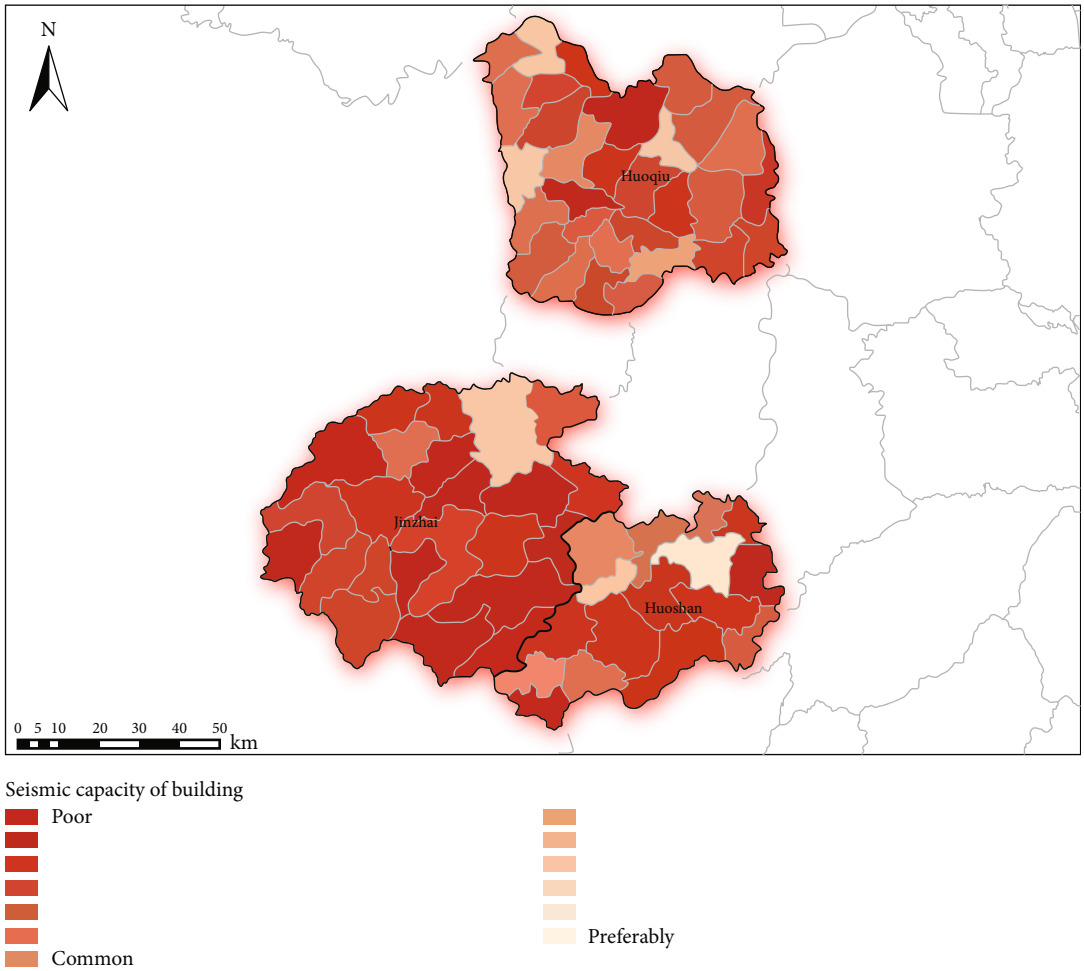

Figure 6: Distribution map of seismic capacity of building areas in Ta-pieh Mountains of Lu'an.

strong earthquakes with a magnitude of 5.0 or above in the region, with the largest earthquake being the Huoshan Luoerling Ms 6.25 earthquake. Over the years, counties and districts near the Ta-pieh Mountains in Lu'an have been designated as earthquake risk areas. Analyzing and researching the seismic capacity of buildings in these areas has practical work significance. The model proposed in this article was used to evaluate the seismic capacity of 455879 buildings in 3 counties, 66 townships, and 455879 buildings in the Ta-pieh Mountains of Lu'an.

The seismic capacity of regional buildings is determined using the seismic performance index of regional buildings [29], and the calculation method is as follows:

$$I = P_{\mathrm{db}} \times I_{\mathrm{db}} + P_{\mathrm{bz}} \times I_{\mathrm{bz}} + P_{\mathrm{yzbz}} \times I_{\mathrm{yzbz}}. \quad (11)$$

I represents the seismic performance index of buildings in townships as a region; P_{db}, P_{bz}, and P_{yzbz} refer to the proportion of buildings in townships with seismic fortification capabilities that meet the standards, are suspected to be insufficient, and are suspected to be severely insufficient; I_{db}, I_{bz}, and I_{yzbz} refer to the seismic performance indices of buildings in townships that meet the seismic capacity standards, are suspected to be insufficient in seismic capacity, and are suspected to be severely insufficient in seismic capacity. They are obtained using expert experience method, with values of 0.18, 0.11, and 0.07, respectively.

Divide the calculated seismic performance index of regional buildings into above 0.18, 0. 09~0. 18, and below 0.09. The three levels respectively, represent good, average, and poor seismic capacity of the house. According to equation (11), calculate the seismic performance index of each township building based on the proportion of seismic fortification types, and determine the seismic capacity level of each township building (Table 7 and Figure 6).

According to the calculation results, except for Meishan Town in Jinzhai County, Luoerling Town in Huoshan County, Hengshan Town in Huoshan County, Zhouji Town in Huoqiu County, Madian Town in Huoqiu County, and Xiadian Town in Huoqiu County, the seismic performance index of all other areas is below 0.09, indicating poor overall seismic capacity. Through sampling surveys, the seismic capacity distribution maps of densely packed building in different urban areas were verified. There are many civil structural buildings in Shuanghe Town, Youfangdian Township,

Taoling Township, Gubei Town, Zhufo'an Town, Taipingfan Township, Foziling Town, Dongxixi Township, Xihu Township, and Huhu Town, with poor seismic capacity, in which other regions have many brick and wood structures, and a certain proportion of brick and concrete structures exist. The overall seismic capacity is common, which is consistent with the calculation results. Based on the evaluation system indicators of seismic capacity of building structures, analysis was conducted on the buildings in various townships. The types of buildings in each region basically include frame structures, standard brick concrete, nonstandard brick concrete, and a small number of old civil structures and brick wood structures. The buildings that are estimated to meet the seismic capacity standards are mainly residential buildings, industrial factories, and commercial buildings with multistory or high-rise frame structures built in urban areas and market towns after 2000. The building materials of the houses are mostly steel bars and brick concrete, and they are designed and constructed according to building standards. Residents have a wide range of knowledge about earthquake prevention and disaster reduction, and there is no unauthorized house renovation or building behavior. Most buildings suspected to have insufficient seismic capacity are old buildings in urban areas, with self-built multistory houses in the urban-rural fringe. One- to two-story buildings are uniformly planned along national roads, provincial roads, county roads, etc. Some of them are designed and constructed according to building codes, and some buildings have been privately renovated. Suspected to have severe seismic capacity, most of them are single-story houses in urban villages and urban-rural fringe areas, as well as 1-2-story low-rise self-built houses with irregular sheeting characteristics distributed on rural roadside and farmland. They have not been designed and constructed according to building standards, and there are many cases of private house renovation and disorderly construction. Residents have weak awareness of earthquake prevention and disaster reduction.

5. Conclusion

(1) A judgment system for seismic capacity of building structures was constructed using FTA method. The seismic capacity of building structures is greatly influenced by indicators such as foundation bearing capacity, foundation bearing capacity, seismic facilities, overall connection structure, building structure, and building seismic design specifications

(2) A sample set was constructed based on the basic data of some houses and buildings in Huoshan County. The K-SMOTE mixed sampling method was used to solve the problem of sample imbalance, and the naive Bayesian algorithm was improved through random crossover validation to construct a model for determining the seismic capacity of houses and buildings. The model integrated the experience of multiple experts in the field of earthquake disaster assessment and obtained the seismic capacity level of houses and buildings through quantitative indicator evaluation, providing a new method for determining the seismic capacity of building structures

(3) This model can be effectively used to quickly determine the seismic capacity of building structures, and compared with traditional naive Bayesian algorithms, the model constructed by the method proposed in this paper has better generalization and higher accuracy

(4) Applying this method to the evaluation of seismic capacity of buildings in the Dabie Mountain area of Lu'an, the seismic performance index of buildings in various townships is obtained, and the overall seismic capacity of buildings in the Dabie Mountain area of Lu'an is poor. Therefore, the implementation and management of seismic fortification measures should be strengthened

Data Availability

The basic data source used in this research institute is the local building property database, and the copyright belongs to the Anhui Provincial Seismological Bureau. If you need to use it, you can contact the author to provide it.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

We would like to express our deep gratitude to Professor Dan Pan, who provided assistance and guidance to the research work of this article. This work was supported by the Earthquake Emergency and Information Youth Fund of China Earthquake Administration under Grant CEAITNS202320.

References

[1] B. Sun and P. Yan, "Damage characteristics and seismic capacity of buildings during Nepal M s 8.1 earthquake," *Earthquake Engineering and Engineering Vibration*, vol. 14, no. 3, pp. 571–578, 2015.

[2] R. Couto, M. V. Requena-García-Cruz, R. Bento, and A. Morales-Esteban, "Seismic capacity and vulnerability assessment considering ageing effects: case study—three local Portuguese RC buildings," *Bulletin of Earthquake Engineering*, vol. 19, pp. 6591–6614, 2021.

[3] R. Montuori, E. Nastri, V. Piluso, and P. Todisco, "Evaluation of the Seismic Capacity of Existing Moment Resisting Frames by a Simplified Approach: Examples and Numerical Application," *Applied Sciences*, vol. 11, no. 6, p. 2594, 2021.

[4] A. Masi and M. Vona, "Vulnerability assessment of gravity-load designed RC buildings: evaluation of seismic capacity through non-linear dynamic analyses," *Engineering Structures*, vol. 45, pp. 257–269, 2012.

[5] M. Yazdi, "Hybrid probabilistic risk assessment using fuzzy FTA and fuzzy AHP in a process industry," *Journal of Failure Analysis and Prevention*, vol. 17, no. 4, pp. 756–764, 2017.

[6] Q. Zhang, Y. Zhuang, Y. Wei, H. Jiang, and H. Yang, "Railway safety risk assessment and control optimization method based on FTA-FPN: a case study of Chinese high-speed railway station," *Journal of Advanced Transportation*, vol. 2020, Article ID 3158468, 11 pages, 2020.

[7] N. Hosseini, S. Givehchi, and R. Maknoon, "Cost-based fire risk assessment in natural gas industry by means of fuzzy FTA and ETA," *Journal of Loss Prevention in the Process Industries*, vol. 63, article 104025, 2019.

[8] M. Zhang, W. Song, Z. Chen, and J. Wang, "Risk assessment for fire and explosion accidents of steel oil tanks using improved AHP based on FTA," *Process Safety Progress*, vol. 35, no. 3, pp. 260–269, 2016.

[9] M. J. Jafari, M. Pouyakian, A. Khanteymoori, and S. M. Hanifi, "Development of a framework for dynamic risk assessment of environmental impacts in chemicals warehouse using CFD-BN," *International Journal of Environmental Science and Technology*, vol. 18, no. 10, pp. 3189–3204, 2021.

[10] S.-S. Leu and C.-M. Chang, "Bayesian-network-based safety risk assessment for steel construction projects," *Accident Analysis & Prevention*, vol. 54, pp. 122–133, 2013.

[11] H. Bian, J. Zhang, R. Li, H. Zhao, X. Wang, and Y. Bai, "Risk analysis of tripping accidents of power grid caused by typical natural hazards based on FTA-BN model," *Natural Hazards*, vol. 106, no. 3, pp. 1771–1795, 2021.

[12] A. Kokangül, U. Polat, and C. Dağsuyu, "A new approximation for risk assessment using the AHP and Fine Kinney methodologies," *Safety Science*, vol. 91, pp. 24–32, 2017.

[13] A. F. Guneri, M. Gul, and S. Ozgurler, "A fuzzy AHP methodology for selection of risk assessment methods in occupational safety," *International Journal of Risk Assessment and Management*, vol. 18, no. 3/4, p. 319, 2015.

[14] H.-M. Lyu, W.-J. Sun, S.-L. Shen, and A.-N. Zhou, "Risk assessment using a new consulting process in fuzzy AHP," *Journal of Construction Engineering and Management*, vol. 146, no. 3, article 04019112, 2020.

[15] Q. Dong and O. Cooper, "An orders-of-magnitude AHP supply chain risk assessment framework," *International Journal of Production Economics*, vol. 182, pp. 144–156, 2016.

[16] Y. Wang, L. Cui, C. Zhang, W. Chen, Y. Xu, and Q. Zhang, "A two-stage seismic damage assessment method for small, dense, and imbalanced buildings in remote sensing images," *Remote Sensing*, vol. 14, no. 4, p. 1012, 2022.

[17] Y. Xu, S. Wei, Y. Bao, and H. Li, "Automatic seismic damage identification of reinforced concrete columns from images by a region-based deep convolutional neural network," *Structural Control and Health Monitoring*, vol. 26, no. 3, article e2313, 2019.

[18] Y. Xu, Y. Li, X. Zheng, X. Zheng, and Q. Zhang, "Computer-vision and machine-learning-based seismic damage assessment of reinforced concrete structures," *Buildings*, vol. 13, no. 5, p. 1258, 2023.

[19] M. Langarizadeh and F. Moghbeli, "Applying naive Bayesian networks to disease prediction: a systematic review," *Acta Informatica Medica*, vol. 24, no. 5, pp. 364–369, 2016.

[20] S. Chen, G. I. Webb, L. Liu, and X. Ma, "A novel selective naive Bayes algorithm," *Knowledge-Based Systems*, vol. 192, article 105361, 2020.

[21] S. Xu, "Bayesian Naïve Bayes classifiers to text classification," *Journal of Information Science*, vol. 44, no. 1, pp. 48–59, 2018.

[22] H. Chen, S. Hu, R. Hua, and X. Zhao, "Improved naive Bayes classification algorithm for traffic risk management," *EURASIP Journal on Advances in Signal Processing*, vol. 2021, no. 1, pp. 1-2, 2021.

[23] G. Arcidiacono, "Development of a FTA versus parts count method model: comparative FTA," *Quality and Reliability Engineering International*, vol. 19, no. 5, pp. 411–424, 2003.

[24] W. Ge, Y. T. Jiao, and X. Q. Hong, "Risk assessment of life loss caused by dam breach based on AHP-BN method," *Journal of Zhengzhou University*, vol. 42, no. 3, pp. 8–12, 2019.

[25] F. Yu, *Linear Algebra and Its Applications*, Tongji University Press, Shanghai, 2014.

[26] R. Blagus and L. Lusa, "SMOTE for high-dimensional class-imbalanced data," *BMC Bioinformatics*, vol. 14, no. 1, p. 106, 2013.

[27] H. B. Shi, Y. W. Chen, and X. Chen, "Summary of research on SMOTE oversampling and its improved algorithms," *CAAI Transactions on Intelligent Systems*, vol. 14, no. 6, pp. 1073–1083, 2019.

[28] C. Wang and K. Wang, "An improved random forest algorithm based on decision trees clustering reduction," *Journal of Nanjing University of Posts and Telecommunications*, vol. 39, no. 3, pp. 91–97, 2019.

[29] X. D. Liu, Z. H. Liu, and L. Liu, "Features and seismic capacity evaluation of rural houses in Zhangjiakou area," *Journal of Seismological Research*, vol. 42, no. 2, pp. 158–165, 2019.