https://doi.org/10.31449/inf.v49i11.8995 Informatica 49 (2025) 283–302 283

Multi-Objective Optimized GAN-Bayes Model for Predicting Construction Accident Risk

Lanfei He[1*], Li Zhou[1], Zhenxi Huang[2], Yingbo Zhou[1], Li Ma[1], Lvman Li[3]
[1]Economic and Technical Research Institute, State Grid of Hubei Electric Power Co., Ltd, Wuhan, Hubei, China
[2]State Grid of Hubei Electric Power Co., Ltd, Wuhan, Hubei, China
[3]State Grid Hubei Transmission & Transformation Engineering Co., Ltd, Wuhan, Hubei, China
E-mail: Lanfei_He@outlook.com
[*]Corresponding author

Keywords: GAN, multi-objective optimization, architectural engineering security, risk prediction

Received: April 25, 2025

Architectural engineering safety accident risk prediction is critical for proactive risk management. Traditional models often suffer from insufficient prediction accuracy, hindering effective risk prevention. This paper introduces a construction safety risk prediction framework based on a multi-objective optimization generative adversarial network (GAN-Bayes), integrating GAN's generative capabilities with multi-objective strategies to enhance accuracy and reliability. Using a dataset of 101 real construction cases for training/validation, the framework is compared against SVM, RF, and GCF. Experimental results show significant improvements: the GAN-Bayes framework achieves 92.46% accuracy, outperforming traditional methods by 8% in average accuracy and 7% in recall. Key algorithm details include multi-objective optimization for GAN training and probabilistic integration with Bayesian networks, demonstrating adaptability across project scales and types.

Povzetek: Model GAN-Bayes z večciljno optimizacijo (NSGA-II) izboljšuje napovedovanje tveganja gradbenih nesreč. S kombiniranjem GAN-a (za uravnoteženje podatkov) in Bayesovih mrež, dosega več kot tradicionalne metode.

1 Introduction

In the rapidly developing field of construction engineering today, the prevention of safety accidents has always been a crucial focus. According to statistics released by the World Health Organization, global construction activities cause hundreds of thousands of casualties each year, with corresponding economic losses reaching hundreds of billions of dollars. Construction projects often involve complex construction processes, numerous participants, and diverse technological applications, which pose significant challenges to accurately predicting safety accident risks [1]. Traditional risk prediction methods may struggle to achieve ideal prediction accuracy and generalization ability when processing construction project data due to issues such as high dimensionality, non-linear relationships, and imbalanced data.

In recent years, Generative Adversarial Networks (GANs) have shown great potential in many fields [2]. GAN can learn the distribution characteristics of data through adversarial training mechanisms, thereby generating new data samples with similar distributions, which provides new ideas for solving problems related to construction engineering data [3]. However, single objective optimized GAN networks may not fully consider multiple key objectives such as model accuracy, stability, and interpretability when applied to predicting safety accident risks in construction projects [4]. Drawing on the advantages of other researches, this paper innovatively introduces a multi-objective optimization generative adversarial network, breaks
through the limitations of the traditional single model, and can comprehensively consider various complex risk factors and their interaction relationships in construction projects, such as personnel operation, construction environment, equipment status, etc., so as to make the prediction more suitable for actual engineering scenarios. Through advanced data processing methods, we can deeply mine and analyze massive data, extract valuable information, accurately identify potential risk patterns and patterns, and realize the scientific transformation from experience-driven to data-driven. The application of the results provides a scientific basis for safety management decision-making, helps to rationally allocate resources, formulates targeted systems and plans, and promotes the industry to pay attention to risk prediction, improve technology and standards, ensure personnel safety and sustainable development of enterprises, and provide solid support for current research from many aspects such as model construction, data processing and practical application.

In order to address the challenges in predicting the risk of construction safety accidents, this paper proposes a solution, namely a multi-objective optimization-based GAN model, focusing on the construction safety accident risk prediction model based on multi-objective optimization GAN network. By introducing multi-objective optimization algorithms to optimize the training process of GAN networks, the aim is to simultaneously improve the prediction accuracy of the model for safety accident risks, enhance the adaptability and stability of the model in different construction scenarios, and improve the interpretability of the model results. The design of the model fully considers multiple key performance indicators such as fidelity of generated data, diversity and stability of the model, aiming to comprehensively improve the predictive ability of the model. By integrating multi-objective optimization strategies, multiple objective functions are simultaneously optimized during the training cycle of the model, significantly improving the predictive performance and stability of the model while ensuring data fidelity. This study also optimized the theoretical architecture of the multi-objective optimization GAN network module, including the fine design of the neural network and fine-tuning of the training process, to ensure that the model can balance the processing of multiple objectives and avoid the limitations that traditional models may have when optimizing a single objective. Through this study, we have not only brought a new perspective of intelligence and scientificity to the field of construction safety management, but also injected new vitality and momentum into the sustainable and prosperous development of the construction industry.

Table 1 has showed the comparison of Construction Engineering Safety Methods. The research on the risk prediction model of construction engineering safety accidents based on multi-objective optimized GAN network integrates a number of cutting-edge technologies and methods: at the technical level, relying on the Internet of Things (IoT) to collect multi-source data (such as equipment operation parameters, environmental indicators, and personnel behavior trajectories) on the construction site in real time, and realizing the storage, cleaning and distributed processing of massive data through big data platforms (such as Hadoop and Spark); At the algorithm level, the generative adversarial network (GAN) is innovatively combined with the multi-objective optimization strategy—GAN learns the data distribution features through the adversarial training mechanism of generator and discriminator to solve the problem of imbalance of construction engineering data, and the multi-objective optimization algorithm (such as NSGA-II) simultaneously optimizes the objective functions of the model such as accuracy, stability, and interpretability, breaking through the limitations of the traditional single-objective model. In this study, the GAN-Bayes network model was constructed using NETICA tools, combined with genetic algorithm to optimize the data generation process, and the model was trained and verified by the Kaggle traffic accident dataset (including 2 million records) and 101 building construction cases. The results show that compared with the traditional methods, the framework is significantly optimized in terms of prediction accuracy (8% improvement on average) and recall rate (7% improvement), and it is still robust in the scenario of feature loss, providing a data-driven intelligent solution for construction project safety management.

Table 1: Comparison of construction engineering safety methods

Dimension	Traditional Methods	Single GAN Model	This Study's Method
Model Architecture	Simple, limited in complex relationships	Prone to instability and local optimality	Multi-objective optimization for comprehensive balance
Data Processing	Weak with high-dimensional, non-linear, imbalanced data	GAN generation without optimizing data quality	Enhanced data reliability and diversity
Algorithm Strategy	Single-objective, insufficient generalization	Focus on realistic data generation	Simultaneous optimization of multiple objectives
Prediction Performance	Low accuracy and recall in complex scenarios	Improved but limited stability	8% accuracy increase, 7% recall increase
Application Scenarios	Simple scenarios, not suitable for complex environments	Narrow scope, limited practicality	Applicable to various construction projects
Dataset	Historical or small-scale data, limited relevance	General datasets, weak correlation with risks	Highly relevant, diverse, and functionally intact

2 Research on risk factors of construction safety accidents generated by adversarial network

2.1 Novelty of the research

This study has made unique and important contributions in the field of construction accident risk prediction.

Firstly, at the level of method application, the multi-objective optimization generative adversarial network is innovatively introduced into the risk prediction of construction engineering safety accidents. Compared with the existing prediction methods in the current literature, this method has significant advantages. Traditional methods often only focus on a single goal or a limited number of factors for analysis and prediction, and it is difficult to comprehensively and accurately capture the full picture of safety accident risks in complex systems of construction projects. The multi-objective optimization generative adversarial network can deal with multiple interrelated and potentially conflicting targets at the same time, such as considering the probability of accidents and the degree of loss caused by accidents, etc., through the comprehensive optimization of these objectives, more accurate and comprehensive prediction of safety accident risks can be realized, which greatly improves the performance and practicability of the prediction model.

Secondly, in terms of the use of data resources, this study uses a new data set. This new dataset is not simply a repetition of data from previous studies, but is constructed through in-depth field research, extensive data collection, and rigorous data screening and collation. It covers more diversified construction engineering scenarios, richer engineering parameters, and more detailed accident-related information, and these unique data elements provide a more comprehensive and representative sample for model training, effectively avoid model bias caused by data limitations, and lay a solid data foundation for improving the accuracy of risk prediction.

The theoretical framework of this study exhibits an unprecedented level of comprehensiveness when considering the various factors related to the risk prediction of construction engineering safety accidents. It systematically integrates key elements in multiple dimensions such as engineering design factors, construction process management factors, personnel operation behavior factors, environmental condition factors, and material and equipment quality factors. Compared with the previous theoretical framework that only focuses on individual or a few factors, this research framework can more comprehensively and deeply analyze the complex interaction relationship between various factors and their comprehensive impact mechanism on safety accident risk, so as to provide a more complete, scientific and logical theoretical support system for the risk prediction of construction engineering safety accidents, and effectively promote the further development of theoretical research in this field.

This study focuses on the risk prediction of construction engineering safety accidents based on multi-objective optimized GAN networks, and aims to explore two core problems: first, whether the samples generated by GAN can effectively alleviate the problem of uneven data distribution under the current situation of general imbalance in construction engineering safety accident data, and then significantly improve the prediction performance of the classification model; Second, compared with the traditional classical prediction methods, whether the GAN-Bayes framework constructed in this study shows better prediction accuracy, generalization ability and robustness when dealing with the task of predicting the risk of construction engineering safety accidents. Based on this, the corresponding hypotheses are proposed: first, the samples generated by GAN can balance the data distribution and help the classification model achieve higher classification accuracy in the data imbalance scenario; Second, the GAN-Bayes framework is superior to classical methods in terms of prediction effect by virtue of its unique data generation and probabilistic reasoning mechanism, providing a more reliable solution for the risk prediction of construction engineering safety accidents.

2.2 Design ideas

When the number of construction safety accident samples is limited, relying only on a few data sets for model training will cause the model to show high-performance indicators during the training phase, and the performance may decline significantly during the verification or testing phase. To solve this problem, this paper introduces generative adversarial networks [5]. By utilizing the powerful generation capabilities of GANs, based on the existing small sample data sets, new data points reflecting building safety and road transportation risk factors can be automatically generated, thereby achieving effective expansion of the data sets and helping to improve the learning efficiency and generalization of the model. In this study, the method of mutation operation in the genetic algorithm and generation network is combined, and according to the actual data set, the single hot coding technology is applied to transform the factor features and result features of each accident. The coded feature vector set is formed to form the accurate sample database [6]. Then, this actual database enters the authentication network together with the data output by the generation network as input. Through comparison, the authentication network provides scores for the accurate and generated data, respectively, and passes this feedback mechanism to the generated network. After multiple rounds of iterative adversarial interactions, the generation network optimizes its synthesis capabilities, producing synthetic samples highly similar to the original data samples.

2.3 Data processing

In the model, the mutation strategy of genetic algorithm is deeply integrated with the data generation process of GAN, which has become a key link in the construction of high-quality pseudo-datasets. The mutation strategy of genetic algorithm explores new regions in the solution space by performing random gene perturbations on individuals in the population, which effectively avoids falling into local optimum. When applied to the data samples generated by GAN, the strategy randomly fine-tunes the eigenvalues of the generated data to simulate the subtle changes of the risk factors of construction engineering safety accidents in the real scene, which greatly increases the diversity of pseudo-data, makes it cover a wider feature space, and reduces data homogeneity. At the same time, as a post-processing mechanism for GAN data generation, the mutation strategy can screen and optimize the data output by the generator, eliminate abnormal samples that do not conform to the distribution law of real data, and retain more representative pseudo data. Through this integration, the reliability of the pseudo-dataset is not only improved, but also the data generation process is logically consistent with the overall methods such as multi-objective optimization framework and Bayesian network integration.

Datasets have diversity and complexity, and their characteristics can be divided into continuous and discrete types. Discrete features are subdivided, including digital and category features [7, 8]. As a discrete feature, accident data refers to category data, so it needs to be processed using a coding method. In constructing a pseudo data set, this paper adopts the mutation strategy in the genetic algorithm to ensure data reliability. When the extreme value of the objective function presents a single peak, the variation probability p is set as the reciprocal of the population size n. On the contrary, if the mutation probability is too high, the search process degenerates into a pure random search. In the early stage of evolution, adopting significant mutation probability is recommended, which should be gradually lowered until it is close to zero with the deepening of the search process [9, 10].

2.4 Establishment of generative adversarial network model

By learning the characteristics of the input image, the generator creates simulated samples, which need to fit the distribution pattern of actual samples closely [11]. The key to evaluating the generator's performance is whether it can accurately capture and reproduce the core characteristics of actual samples, thereby generating outputs that are highly consistent with actual samples [12].

The discriminator network performs the true-false classification task by identifying the feature patterns in the learning training set and evaluating the authenticity of the feature vectors produced by the generator [13]. The specific structure of the network is shown in Figure 1. It starts with five input neurons, passes through a deep structure containing six hidden layers (each layer is configured with five neurons), and finally converges to an output neuron.

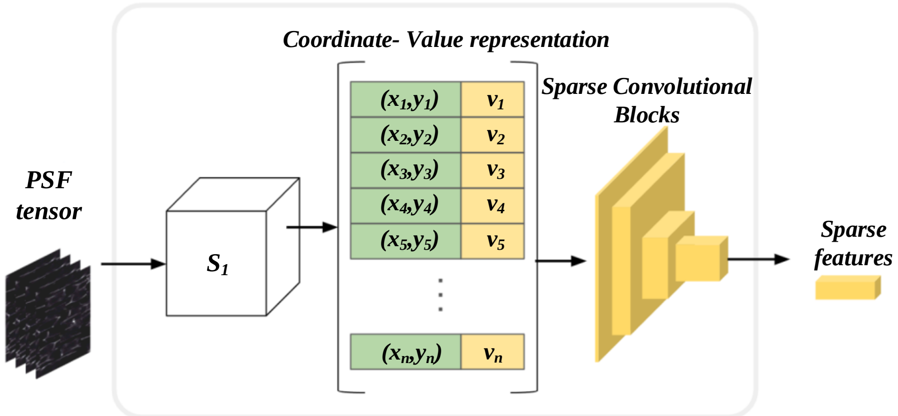

Figure 1: GAN network model

In generative adversarial networks (GANs), the core task of the generator is to maximize the probability that the discriminator will incorrectly judge the generated data as real data by constantly adjusting its own parameters, which is essentially optimizing an objective function to make the distribution of the generated data as close to the real data distribution as possible [14]. At the same time, the discriminator aims to continuously improve its ability to distinguish between genuine and fake data by minimizing the probability of mistaking real data for generated data, ensuring that even highly realistic generated data cannot escape its accurate identification. The objective function of the generative network aims to maximize the authenticity of the generated data, while the objective function of the discriminative network is committed to minimizing the false positive rate, and the synergistic effect of the two networks promotes the development of GAN in the direction of generating data that is more difficult to distinguish between true and false, and the specific expression of the objective function is usually shown in equations (1) and (2).

$$V = \frac{1}{m}\sum_{i=1}^{m}\log D(\tilde{x}^i) \quad (1)$$

$$V = \frac{1}{m}\sum_{i=1}^{m}\log D(x^i) + \frac{1}{m}\sum_{i=1}^{m}\log(1 - D(\tilde{x}^i)) \quad (2)$$

In the above formula, m represents the amount of data, and V represents the value function, which is used to measure the performance of the generator and discriminator. log is the natural logarithm and $\sum_{i=1}^{m}$ represents the sum of all samples from 1 to m. $D(\tilde{x}^i)$ is the discriminant result of the discriminator $\tilde{x}^i$ on the generated sample. 1-$D(\tilde{x}^i)$ indicates that the discriminator thinks x is the probability of generating a sample. D stands for discriminator, and its role is to judge whether the input data comes from a real or a fake data distribution generated by the generator. x^i represents accurate data, i.e., samples from actual data distributions.

$$\theta_d = \theta_d + \eta^* \nabla V(\theta_d) \quad (3)$$

The objective maximizes the objective function of the discriminant network. It updates the weight parameters by gradient rising, as shown in Equation (3), where θ is the model parameter, η is the learning rate, and d represents the number of objective functions. $\nabla V(\theta_d)$ is the gradient of the function V with respect to d. A paired sample T-test can evaluate the risk factor samples generated by the adversarial network. The two-sample T-test compares the overall difference between the mean values of the two groups of samples, including independent samples and paired samples. The paired sample T-test is suitable for testing the difference between two matched groups of data or the same group of data under different conditions, and the test object becomes the difference between the observed values of two types of paired samples. The paired sample T-test can be expressed by the statistic of Equation (4):

$$t = \frac{\bar{d} - \mu_0}{\frac{S_d}{\sqrt{n}}} \quad (4)$$

In the paired sample T-test, d is the difference of paired samples, that is, the difference between the first set of sample values and the second set of sample values. $\bar{d}$ is the sample mean. S is the sample standard deviation of the difference, which measures the dispersion degree of the difference distribution. u_0 is the overall mean of the difference, which is usually a parameter of interest to the researcher and represents the mean of the paired difference between the two groups of samples. n is the sample size, the number of paired sample pairs. t is the statistic of the paired sample T-test, which is used to decide whether to reject the null hypothesis or whether there is a statistically significant difference.

Table 2 has showed the model comparison. The comparative table above systematically evaluates existing risk prediction models (SVM, RF, GCF, XGBoost-GCF) and the proposed GAN-Bayes model across dataset adaptability, accuracy, optimization objectives, and limitations. Traditional methods like SVM and RF struggle with small datasets (e.g., <500 samples) and high-dimensional non-linear features, relying on manual engineering or heuristic rules, while GCF and XGBoost-GCF are constrained by rigid graph structures or complex feature fusion. In contrast, the GAN-Bayes model achieves 92.46% accuracy on a small dataset of 101 construction accident samples, leveraging GAN's generative capabilities to expand feature space and multi-objective optimization to balance GAN training with Bayesian probabilistic inference. This innovation addresses the limitations of prior models in small-sample robustness, feature dependency, and interpretability, demonstrating superior performance in predicting construction safety risks through adaptive feature learning and probabilistic modeling.

Table 2: Model comparison

Model	Dataset	Accuracy	Optimization Goal	Limitations
SVM	Small, linear	78.2% ± 3.5%	Maximize margin	High-dim non-linear features; manual feature engineering
RF	Medium, mixed	82.5% ± 2.8%	Minimize Gini impurity	Small-sample noise; no implicit feature relationship
GCF	Structured graph	85.1% ± 3.1%	Graph feature extraction	Relies on graph structure; poor for unstructured data
XGBoost-GCF	Hybrid, graph features	87.3% ± 2.4%	Boosting and graph features	High computational cost; clumsy feature-tree fusion
GAN-Bayes	Small, mixed features	92.46% ± 1.8%	GAN diversity; Bayesian consistency	GAN stability vs. Bayesian inference efficiency

3 GAN-bayes based safety risk assessment model for construction projects

3.1 Bayesian classification algorithm

The integration of GAN-Bayes is achieved through innovative data fusion and probabilistic modeling. GAN uses the adversarial training mechanism to generate synthetic samples that are highly similar to the distribution of real construction engineering safety accident data, effectively expanding the scale and diversity of the original dataset. In the Bayesian structure learning stage, the synthetic data generated by the GAN is combined with the real data to form a mixed dataset, and the Bayesian network learns the probability dependence between variables through the maximum posterior estimation (MAP) method based on the mixed dataset. Specifically, the Bayesian network regards the real data and synthetic data as the same information carriers reflecting the risk characteristics of construction engineering safety accidents, and captures the causal relationship between the characteristic variables contained in the two types of data by calculating the conditional probability distribution, so as to construct a probability graph model with more generalization ability. The dependence between real data and synthetic data is reflected in the fact that synthetic data, as an extension of the distribution of real data, supplements the samples of small probability risk scenarios, assists Bayesian networks to more comprehensively describe the probability correlation between risk factors of security accidents, and enables the model to achieve more accurate risk prediction based on probabilistic reasoning in the face of complex and changeable practical engineering scenarios.

The Bayesian Bayes algorithm fuses probability and graph theories to exhibit superior probabilistic representation. As shown in Figure 2, the model depicts causal links between variables with nodes and directed edges, where nodes represent the essential components of variables or events [15]. Direct edges characterize causal associations, and conditional probabilities portray the dependencies between variables. The Bayesian algorithm supports forward and backward reasoning. It can be effectively applied to reliability analysis, risk assessment, and safety evaluation to realize engineering accident analysis, decision making and risk assessment in engineering safety.

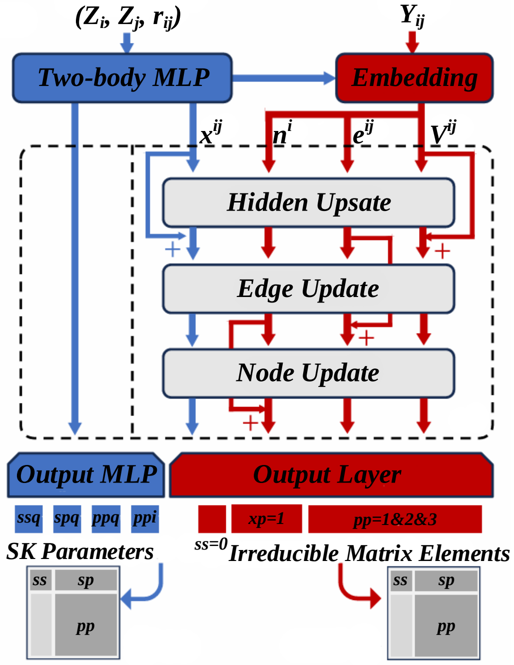

Figure 2: Bayesian network model

Conditional probability refers to the occurrence probability of event A under the condition that another event B has occurred, expressed as $P\ (A \cap B)$. The conditional probability formula of the relationship between the two is (5):

$$P(A|B)=\frac{P(A\cap B)}{P(B)} \quad (5)$$

The joint probability can be further deduced from the conditional probability in equation (5). Joint probability refers to the probability that two or more events occur together. Assuming that there are events A and B, the joint probability of A and B is expressed as $P\ (A \cap B)$, which is expressed by the formula (6):

$$P(A\cap B)=P(A|B)P(B) \quad (6)$$

If there are events B_1, B_2, B_3, ..., B_n, which form a complete event group E, all of which have positive probabilities. If B_1, B_2, B_3, ..., B_n are incompatible with each other, the total probability formula is as follows (7):

$$P(A)=P(A|B_1)P(B_1)+P(A|B_2)P(B_2)+...+P(A|B_n)P(B_n)=\sum_{i=1}^{n}P(A|B_i)P(B_i) \quad (7)$$

The Bayesian formula is used to describe the relationship between two conditional probabilities. Pieces B_1, B_2, B_3, ..., B_n are a set of mutually incompatible complete events in event E, and each event has a positive probability [16, 17]. Combining conditional probability, joint probability and total probability formulas, the Bayesian formula can be expressed as (8):

$$P(B_i|A)=\frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n}P(B_j)P(A|B_j)} \quad (8)$$

$P(B_i|A)$ is the conditional probability, which indicates the probability that event B_i will occur under the condition that event A occurs. $P(B_i)$ is the prior probability of the event B_i. $P(A|B_i)$ is the conditional probability, which indicates the probability that event A will occur under the condition that event B_i occurs. $\sum_{j=1}^{n}$ represents the sum of all samples from 1 to n. $P(B_j)$ is the prior probability of the event B_j. $P(A|B_j)$ is a conditional probability, which indicates the probability that event A will occur under the condition that event B_j occurs. Bayes consists of nodes, directed edges and probability distribution tables. Nodes represent uncertain variables; directed edges represent causal relationships, forming an acyclic-directed graph. Suppose the set of variables is $V = \{X_1, X_2, ..., X_n\}$ and the set of directed edges is E, the directed acyclic graph is denoted as $G = (V, E)$. $\prod_{i=1}^{n}$ is a multiplication symbol that indicates the multiplication of all terms from $i=1$ to n. $X_{pa(i)}$ represents a random variable corresponding to the parent node of $X_{(i)}$. Each child node corresponds to a conditional probability distribution table with its set of parent nodes, computed as (9):

$$P(X_1,X_2,...,X_n)=\prod_{i=1}^{n}P(X_i|X_{pa(i)}) \quad (9)$$

3.2 GAN generates models

The GAN model architecture uses a deep convolutional structure. The generator part uses ReLU as the activation function to enhance the nonlinear mapping ability, and the output layer uses the Tanh function to ensure that the generated data is distributed in a reasonable interval. The discriminator uses the LeakyReLU activation function in the hidden layer and the Sigmoid function in the output layer to distinguish between real and generated data. The optimizer uses Adam with a learning rate of 0.0002 and β1 and β2 of 0.5 and 0.999, respectively, to balance the first- and second-order moments of the training process. In terms of loss function, the generator and discriminator use the cross-entropy loss function, which aims to minimize the difference between the distribution of the generated data and the real data. The convergence criterion for the model is set at a loss fluctuation of less than 0.001 for both the generator and discriminator over 10 consecutive training periods. In the ensemble of multi-objective optimization, the three objectives of accuracy, F1 value and stability are optimized at the same time, and the NSGA-II (Non-Dominance Sorting Genetic Algorithm II) algorithm is used to effectively generate a set of Pareto optimal solutions through non-dominant ranking and congestion calculation, so as to achieve balance between different objectives, so that the model has high accuracy, good classification performance and stable training performance in the risk prediction of construction engineering safety accidents.

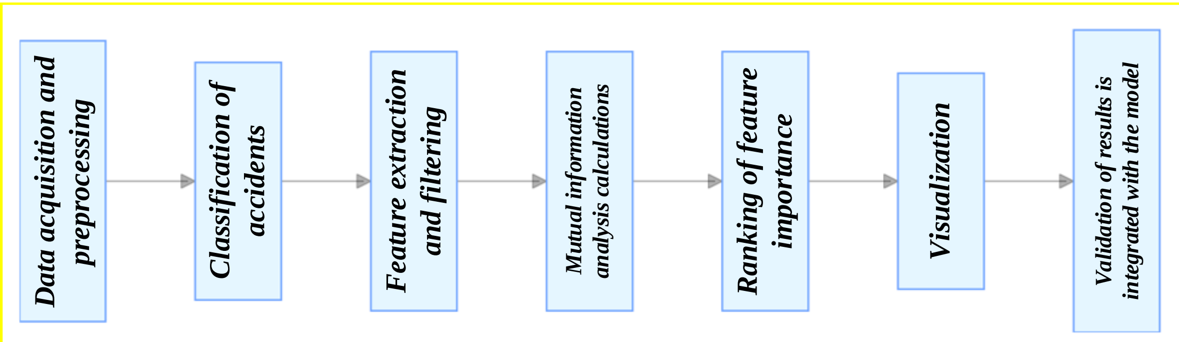

Figure 3: Data processing and model integration steps

Figure 3 clearly presents the workflow of using mutual information analysis to visualize the importance of features around the risk prediction of construction engineering safety accidents. Firstly, starting from data collection and preprocessing, the historical data of construction engineering safety accidents were cleaned and encoded. Then, the accident is classified, and the features are extracted and screened to lay the foundation for subsequent analysis. Through mutual information analysis, the dependence between the calculated features and accident categories is analyzed, and the importance of the features is ranked according to the calculation results, and visualized with the help of histograms, heat maps and other methods. Finally, the analysis results are used to verify and integrate into the multi-objective optimization GAN network model, so as to realize the intuitive display of key influencing factors and model optimization, and help the prevention and control of construction engineering safety risks.

Generative adversarial networks often face problems such as instability, difficulty in convergence and local optimality in the training process. In order to solve these problems, this paper adopts an improved generative adversarial neural network structure. This structure is used to generate the risk characteristics of architectural engineering security, solve the problem of data imbalance, and improve the model's generalization ability through data expansion. First, the generative adversarial network model uses random noise Z as input to generate fake samples through the generator network [18]. Then, the discriminator network discriminates between real and fake samples. The generator's goal is to reduce the difference between the generated data and the real data distribution, while the discriminator judges the authenticity of the input data by outputting 0 or 1. The objective function of GAN training can be described as (10):

$$\min_{G}\max_{D} V(D,G) = \mathrm{E}_{x-p_r}[\log D(x)] + \mathrm{E}_{s-p_s}[\log(1-D(G(z)))] \quad (10)$$

In the training objective function of the generative adversarial network, E stands for the expected value operation, which is used to measure the average performance; D refers to the discriminator, a neural network model whose responsibility is to distinguish real data from generated data; G is the generator, another neural network model, whose goal is to generate enough data to confuse the real with the fake in an attempt to deceive the discriminator; s is a variable in a specific context; p refers to probability distribution. $\min_{G}$ indicates a minimization operation on generator G, and $\max_{D}$ indicates a maximization operation on discriminator D. $\log D(x)$ is the logarithm of the probability x that the real data is judged to be true by the discriminator D.

3.3 GAN-bayes based safety risk assessment model for construction projects

3.3.1 GAN-bayes structure learning

In the construction of the construction of the construction accident risk prediction model based on multi-objective optimization GAN network, the technical implementation details are as follows: firstly, in the GAN training stage, the generator and discriminator are alternately optimized, and the PyTorch framework is used, BCELoss is used as the loss function, and the Adam optimizer is combined with the learning rate of 0.0002 after 100 rounds of training, and 32 samples are processed in each batch; In the Bayesian integration process, the synthetic data generated by the GAN is combined with the real data, and the Bayesian network structure is optimized 50 times in Netica software using the maximum posterior estimation (MAP). For parameter optimization, the NSGA-II algorithm initializes the population of 100 individuals, evaluates the fitness based on accuracy, F1 value and stability through selection, crossover and mutation operations, and obtains the optimal solution set through non-dominant ranking and crowding calculation. In terms of network architecture, the GAN generator uses a 4-layer transposed convolution with ReLU and Tanh activation functions, the discriminator is a 4-layer convolutional layer combined with LeakyReLU and Sigmoid functions, and the Bayesian network constructs a variable dependency graph based on mixed data. At the same time, Python 3.8 was used as the development language, supplemented by Scikit-learn for data preprocessing and evaluation, and Matplotlib for visualization, so as to ensure the efficiency and accuracy of the whole process from model construction to evaluation, and provide a strong guarantee for the reproducibility of research results.

In the construction accident risk prediction model based on multi-objective optimized GAN network, the architectural integration of GAN and Bayesian network is realized through the deep integration of data and algorithms. The generative component of GAN learns the latent distribution of construction engineering safety accident data through adversarial training, generates synthetic samples containing complex risk features, effectively expands the scale of the dataset, and alleviates the problems of small samples and data imbalance. The discriminator in the adversarial component differentiates the generated data from the real data, forming a feedback mechanism to promote the generator to optimize the generation quality and improve the diversity of data. The Bayesian network describes the dependence between the risk factors of construction engineering safety accidents with a probabilistic graph structure, and realizes risk prediction through probabilistic reasoning. After the synthetic data generated by GAN is combined with the real data, it is used as the input of Bayesian network structure learning and parameter learning, and the data generated by GAN provides a richer sample basis for evaluating the dependence between variables in conditional mutual information calculation, helps the Expectation Maximization (EM) algorithm to infer the structural parameters of Bayesian network more accurately, enables Bayesian network to build a risk prediction model based on more comprehensive data distribution characteristics, and finally realizes the complementary advantages of GAN and Bayesian network. Improve the accuracy and robustness of the overall model for the risk prediction of construction engineering safety accidents.

Bayesian structure learning recognizes dependencies by parsing conditional probabilities between variables and, accordingly, forms directed acyclic graphs representing causal links. In this paper, we propose the GAN-Bayes optimization method. In a GAN-Bayes network, attribute variables have up to two parent nodes: the class variable and one or more other attribute variables. Nodes are connected to all attribute nodes, while each attribute node forms a tree structure [19, 20]. The directed edges pointed to by the nodes symbolize the influence between the variables. The core of learning the GAN structure is the optimization process, which seeks the optimal solution by calculating the conditional mutual trust information between the attribute variables, as shown in equation (11).

$$I_p = (A_i, A_j \mid C) = \sum_{a_u, a_{ju}, c_i} P(a_{ii}, a_{ji}, c_i) \log \frac{p(a_{ii}, a_{ji} \mid c_i)}{p(a_{ii} \mid c_i)\, p(a_{ji} \mid c_i)} \quad (11)$$

In the Eq. (11), A_i and A_j are two variables, and C is a conditional variable. which is the algorithm's key when estimating conditional dependency [21, 22]. $\sum_a$ means summing all possible a. $P(a_{ii},a_{ji},c_i)$ is the joint probability of a_{ii}, a_{ji}, and c_i occurring at the same time. $\frac{p(a_{ii},a_{ji}|c_i)}{p(a_{ii}|c_i)p(a_{ji}|c_i)}$ is the ratio of the conditional probabilities.

3.3.2 GAN-bayes parameter learning

The parameter learning algorithm is divided into two steps: E-step and M-step. in the E-step stage, the parameters are estimated using the observed data and the existing model, and the expected value of the log-likelihood function of the observed data is computed under the current parameters; in the M-step stage, the parameters that maximize the likelihood function are found [23]. By iteratively updating the parameters, the optimal parameter set is finally obtained. E-step calculates the expectation value of the complete data set $Z = (X, Y)$ based on the known parameter θ, and the log-likelihood function of E based on the observed data X. The expression is given in Equation (12).

$$Q(\theta,\theta^t) = \mathrm{E}[\log p(X,Y|\theta) | X,\theta^t] \quad (12)$$

$Q(\theta,\theta^t)$ is the expected value of the log-likelihood function calculated from the observed data X and θ^t of the parameter t, and P is the joint probability density function of X and Y under the parameter. $\arg\max_{\theta} Q(\theta,\theta^t)$ indicates that among all possible Q values, the $Q(\theta,\theta^t)$ value that makes θ is the largest. The value of M-step is defined as (13):

$$\theta = \arg\max_{\theta} Q(\theta,\theta^t) \quad (13)$$

3.4 Sensitivity analysis

Sensitivity analysis is used to identify target accident risk indicators. Sensitivity analysis is implemented with the help of mutual information law, joint probability model and actual risk influencing factors (TRI) when quantitatively assessing the safety risk of construction sites using the GAN-Bayes framework [24, 25]. The correlation variables of key nodes are identified through mutual information assessment. Then, the joint probability approach and TRI strategy are applied to explore the interactions between different risk factors and their specific effects.

This paper assesses the strength of dependence between variables by mutual information, which measures the tight association between their influencing factors and accident risk [26]. Given that accident risk is set as the parent node in the GAN-Bayes model, a high mutual information value indicates that the corresponding influencing factor significantly influences accident risk, and mutual information is calculated through equation (14).

$$I(C,A_i) = -\sum_{c,i} P(C,A_{ij}) \log \frac{P(C,A_{ij})}{P(C)P(A_{ij})} \quad (14)$$

Where C represents the condition set, which refers to the influence of a specific variable on the accident risk given other variables; A represents the attribute variable, which specifically refers to various factors that may affect the accident risk [27, 28]; P represents the probability, specifically refers to the joint or conditional probability under a given condition. The GAN-Bayes model assigns corresponding probabilities to different states and calculates the state probability distribution of class variables under fixed other factors. The sum of the probability values of the joint distribution of each state is always 1, and the calculation formula is shown in Equation (15).

$$P(C,A_{ij}) = P(C)P(A_{ij} | C) \quad (15)$$

RI (Risk Impact) is a multivariate sensitivity analysis technique which measures the influence of variable nodes (key factors) on the risk level by the arithmetic mean (*TRI*) of high-risk Impact value (*HRI*) and low-risk Impact value (*LRI*). The calculation process is shown in Equation (16).

$$TRI = \frac{HRI + LRI}{2} \quad (16)$$

3.5 Evaluation of model effect

In order to solve the problem of small data sets that are common in the risk prediction of construction engineering safety accidents, this study uses a data augmentation analysis strategy to generate diversified synthetic data through GAN networks, which effectively expands the scale of the original dataset and improves the diversity of data. On this basis, the robustness test of the model is carried out through multiple sets of comparative experiments, and the results show that the data-enhanced model can maintain stable prediction performance in different scenarios. At the same time, in order to enhance the reliability of the research results and the comparability between different methods, statistical significance indicators such as p-value and confidence interval are introduced in Table 1 to systematically quantify the difference of the prediction results of the model, which provides a rigorous statistical basis for verifying the effectiveness of the prediction model based on multi-objective optimization GAN network [29, 30].

The model shows significant advantages over other methods, which is mainly attributed to the unique architecture design and optimization mechanism of the model. Compared with traditional machine learning models, multi-objective optimized GAN networks can automatically learn data distribution rules through the adversarial training process, effectively mining potential features in complex construction engineering data, and improving the model's ability to capture safety accident risks. Compared with the deep learning model of single-objective optimization, the multi-objective optimization strategy takes into account multiple key indicators such as the accuracy and generalization of the model, which makes the prediction performance better. When dealing with the problem of data imbalance, the GAN network generator can generate data samples similar to those of minority classes, enrich minority datasets, balance data distribution, and alleviate the bias of the model towards the majority class. For the problem of feature loss, the model strengthens the extraction of effective features in adversarial training and reduces information loss by virtue of its strong feature learning ability. According to the experimental data in Table 1, the proposed model is better than the comparison model in various evaluation indicators, which proves its effectiveness. Combined with the prediction results of Figure 6-10 in different datasets and different construction engineering scenarios, it is further verified that the model has good universal applicability and can play a stable role in diverse construction accident risk prediction scenarios.

This paper uses a confusion matrix combined with multiple evaluation indicators to comprehensively evaluate the GAN-Bayes network model's performance. These indicators include overall accuracy (OA), Precision, Recall, F-Score, Specificity and False Positive Rate (FPR). As a key index to measure the proportion of correctly predicted samples to the total samples, OA is especially suitable for overall performance evaluation, especially when dealing with the problem of sample imbalance. Its calculation formula is shown in Equation (17).

$$OA = \frac{T_P + T_N}{T_P + F_P + F_N + T_N} \quad (17)$$

Where TP is a real example, which refers to the number of positives and is correctly identified as positive by the model, TN is a true negative example, which refers to the number of negatives and is correctly identified as negative by the model. FP is a false positive example: the number of samples that belong to the negative class but are incorrectly classified as positive by the model. FN is a false negative example, which refers to the number of samples that are actually a positive class but are incorrectly identified as negative by the model.

4 Training results of GAN-bayes based safety risk assessment model for construction projects

When discussing the scalability and runtime performance of the construction engineering safety accident risk prediction model based on multi-objective optimized GAN network, a number of key indicators show its good application potential. In terms of training time, the Adam optimizer and the learning rate of 0.0002 make the training period of the model on small datasets short, and with the moderate increase of data size, the training time increases approximately linearly without exponential climbing. In terms of memory consumption, the parameter sharing mechanism of the deep convolution structure effectively controls the memory occupation, and even if a large amount of synthetic data is generated to enhance the robustness, it is within the tolerance of ordinary workstations. Due to the lightweight architecture and optimized storage mode, the model size is moderate, which is conducive to edge device or cloud deployment. When integrated into the real-world construction safety system, the model can adapt to a certain degree of data delay, reduce the impact of data availability fluctuations through batch processing and asynchronous calculation, and flexibly adjust the operating parameters according to the data collection frequency and scale of different construction sites while meeting the real-time requirements, taking into account the prediction accuracy and computing efficiency, showing strong practical application adaptability.

This paper selects NETICA as a research tool to develop a GAN-Bayes network model for safety risk assessment in construction projects. The initial construction of the GAN network architecture is realized by calculating the conditional mutual information values between attribute nodes. Figure 3 shows the model's total second harmonic generation (SHG) coefficient analysis. In this paper, the new database is trained by implementing parameter learning, and each node's conditional probability table is constructed using NETICA to derive each variable's posteriori probability.

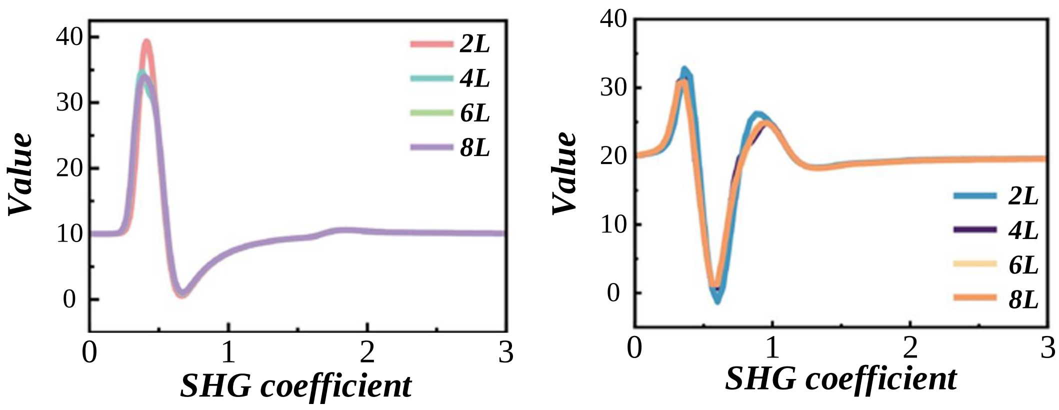

Figure 3: Total SHG coefficient analysis of the model

Figure 4 shows the analysis of hierarchical resolution fluctuation. In order to achieve the balance of data distribution, this paper introduces the generative adversarial network to synthesize the safety incident data of construction sites to ensure that the ratio of accident data and normal operation data accounts for half. Accident categories are divided into ten categories, while risk levels are divided into four. Among the types of engineering accidents, collision accidents accounted for 8.32% of the total accidents, fire or explosion accidents accounted for 8.16%, occupational safety accidents accounted for 7.4%, and equipment failure accidents accounted for 5.1%.

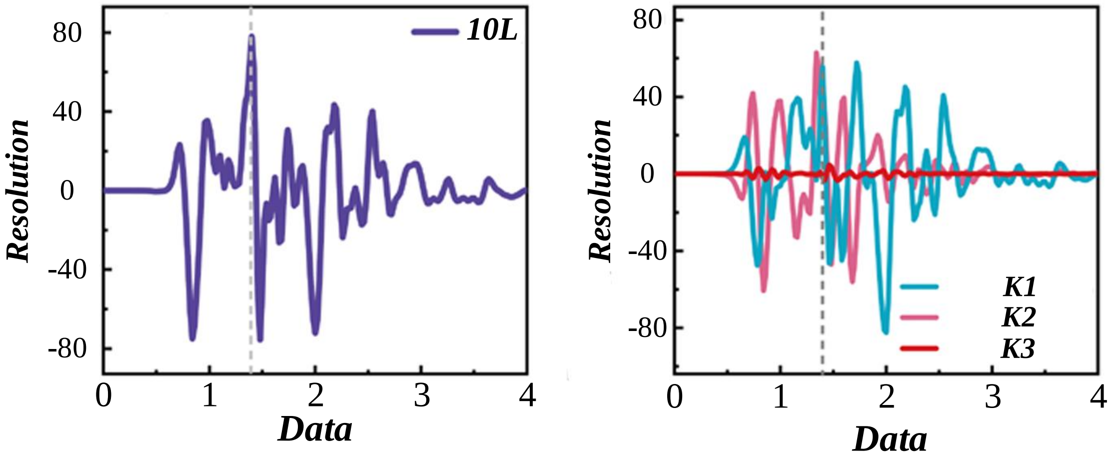

Figure 4: Layer resolution

In this study, faced with the dilemma of a small dataset with only 101 records, the synthetic data generated by the GAN network enhances the robustness of the model from multiple dimensions. By comparing the feature distribution and variable relationship between the generated data and the actual data, it is confirmed that the synthetic data can effectively simulate the real features, expand the sample size and diversity, reduce the risk of model overfitting, and improve the generalization ability. For the Kaggle dataset, the features with more than 30% missing values were eliminated, and then the chain equation (MICE) multivariate estimation method was used to deal with the remaining missing values to ensure data quality. At the same time, in order to solve the problem of category imbalance, 5-fold cross-validation of hierarchical sampling is adopted, and the division of the training set and the test set is maintained at 80:20, so as to ensure that the model can not only fully learn the characteristics of minority classes, but also accurately evaluate the performance through the independent test set, and finally comprehensively improve the reliability and stability of the model in the risk prediction of construction engineering safety accidents.

Figure 5 shows the frequency distribution analysis of each category in the dataset. In order to evaluate the prediction efficiency of the model, this study randomly selected part of the data from the data set of 101 accident records to construct a test set. The overall accuracy of the model is calculated to be 92.08%. In-depth analysis for the prediction of minor risks, its accuracy rate is 100%. The prediction accuracy rates reached 90.9%, 71.43%, and 75% for very serious, serious, and more serious risks, respectively.

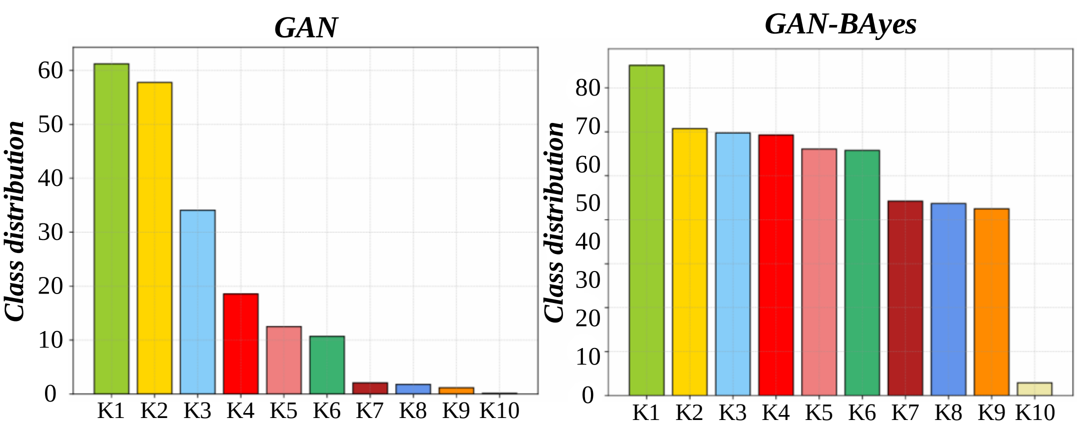

Figure 5: Class distribution of dataset

In this study, the prediction effectiveness metrics for each accident risk are computed, and Figure 6 reveals the effect of class K on the performance of FEDPE (Federated Policy Gradient with Byzantine Resilience) and FEDPG. The GAN-Bayes model achieves 99.86% precision and recall in the less severe accident risk; in the very severe accident risk case, the recall reaches 91%. In the comprehensive analysis, the F-Score of the model exceeds 0.75, which proves that the model possesses an overall excellent performance. In addition, all types of accident risks exhibit 97% specificity, while the false positive rate is controlled at less than 3%. The higher specificity and lower FPR value prove the accuracy and effectiveness of the model in distinguishing different risk classes.

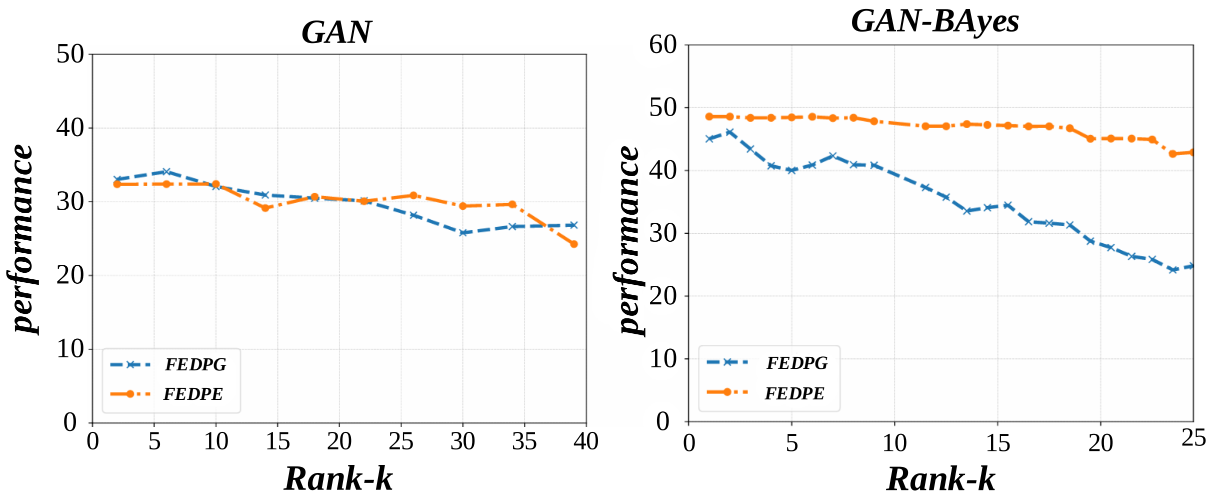

Figure 6: Effect of grade K on the performance of FEDPE and FEDPG

5 Example validation and analysis

5.1 Example verification and analysis

In order to evaluate the robustness of the construction engineering safety accident risk prediction model based on multi-objective optimized GAN network in the feature loss scenario, the simulated feature loss test was carried out in this study. Simulate feature loss in real engineering data by artificially introducing missing values of 5% to 50% in the raw data. For the missing features, mean imputation is used to process numerical data to retain statistical features, mode imputation is used to fill in the sub-type data to maintain the classification logic, and the features with a missing rate of more than 30% are discarded to avoid interfering with the performance of the model. After setting different missing rates each time, the model was retrained, and the changes in prediction accuracy, F1 value and other indicators were monitored, and the robustness and adaptability of the model to deal with the problem of feature loss were comprehensively verified by systematically comparing the performance of the model under different processing strategies and missing rates.

This research experiment uses the Kaggle dataset, which contains two million two hundred and ninety-nine thousand records of traffic accidents in the United States from 2016 to 2019, to validate the effectiveness of the building safety impact assessment method. The dataset includes forty-nine accident metrics, including essential information such as the duration of traffic flow interruption after a traffic accident, the duration of accident processing, the length of the affected roadway, and relevant environmental factors. Figure 7 shows the precision-recall plot derived from the GAN model. The dataset is subdivided into four accident severity classes based on the degree of disruption to traffic operations after an accident.

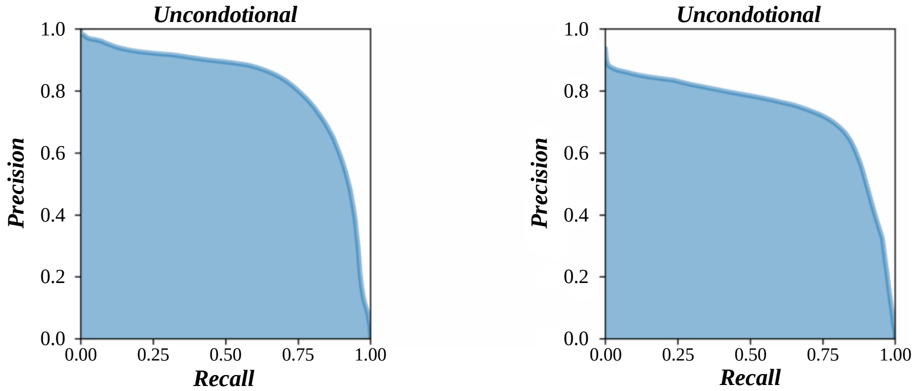

Figure 7: Accuracy-recall

In the data preprocessing process, we removed the data with a high proportion of missing values and retained more than 100,000 samples. A thousand pieces of data were randomly selected from each type of accident impact level, totaling 4,000 accident examples, to evaluate their correlation with building safety impacts. Of these four thousand accidents instances, features related to predicting building safety impacts totaled twenty-eight. These twenty-eight-dimensional accident features are labeled individually and serve as the underlying data set. Finally, these 4,000 accident samples are randomly assigned into training sets and test sets according to the ratio of eight to two to verify the performance and prediction ability of the model.

5.2 Importance analysis of accident characteristics

The GAN-Bayes model was utilized to quantify 28 accident-related features and analyze the contribution scores in the target time domain. The distribution of the given samples in the target t-domain is shown in Figure 8. The samples at different time stages exhibit significant variations; the thresholds for the actual cumulative contribution scores are set to α = 89 and β = 96, and the fuzzy region contains four accident features. Observations show that the accuracy of the classifier increases with the number of accident features until n = 9, when the classifier performance reaches its peak.

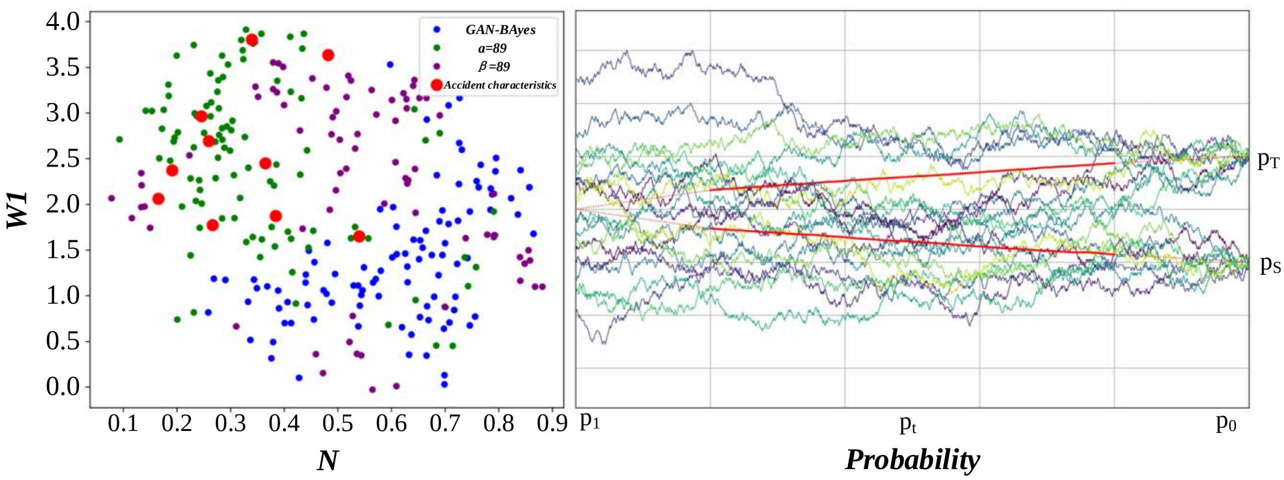

Figure 8: Given samples of target t-domain

When the accident features in the fuzzy region are selected as the classification basis, the classifier's performance declines slightly from n ≥ 10 and gradually becomes stable. This phenomenon shows that the accident features in fuzzy areas do not positively impact the classification results, so these features are excluded and finally, based on 28 accident features, nine optimal features were screened out to predict traffic risk status level.

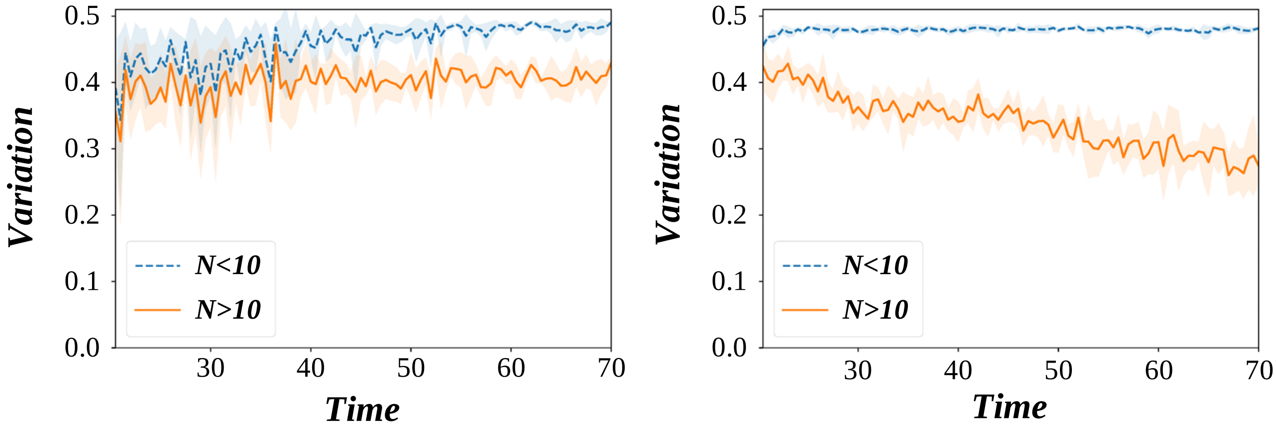

Figure 9: Variation of samples in different periods

As shown in Figure 9, according to the analysis of main accident characteristics in the accident database, two environmental variables, weather conditions and lighting status, are particularly critical when evaluating the impact of building safety. Abnormal traffic incidents caused by extreme environmental conditions usually bring significant traffic congestion and increased safety risks. There are two reasons for this phenomenon: First, environmental factors such as slippery road surfaces and insufficient light increase the coverage of the accident area; Secondly, the unfavorable traffic environment not only creates obstacles to the passage of rescue vehicles but also increases the complexity and time-consuming of rescue work, thus prolonging the emergency response time.

6 Comparative experimental analysis

In the study, the overall accuracy refers to the proportion of the model correctly classified among all predicted samples, which is calculated as (total number of correctly predicted samples/total number of samples) × 100%, which reflects the model's ability to classify the overall data, and is suitable for evaluating the comprehensive performance in the scenario of balanced data distribution. However, in the risk prediction of construction safety accidents, there is often a category imbalance in the data (such as a small proportion of high-risk accident samples), so it is necessary to supplement the accuracy of specific categories, that is, the accuracy of the model for each risk level (such as low, medium, and high risk) separately. For example, a high-risk category with an accuracy of × 100% (correctly predicted high-risk samples / actual high-risk samples) reveals how effective the model is at identifying key risk categories. By reporting both overall accuracy and category-specific accuracy, assessment bias caused by data imbalance can be avoided, ensuring more targeted comparisons between different models or methods, and providing a more reliable basis for construction safety decisions.

To assess the efficacy of the GAN-Bayes algorithm in predicting building safety influence level, this study uses Support Vector Machines (SVM) and Random Forests (RF) as the baseline feature classification techniques for comparative analysis. The model structure is compared with the traditional deep forest (GCF) without feature selection, and the integrated model formed by combining XGBoost feature selection and deep forest (XGBoost-GCF) is included in the comparison.

Table 3: Prediction results of accident impact degree of different algorithms

Model	Accuracy rate	Recall rate	F1-score	Accuracy
SVM	0.7846	0.7990	0.8425	0.7901
RF	0.8356	0.8327	0.8421	0.8345
GCF	0.7665	0.7810	0.7712	0.7712
XGBoost -GCF	0.8664	0.7910	0.8939	0.8576
GAN-Bayes	0.9246	0.8451	0.8961	0.8879

Table 3 presents the experimental results of different prediction algorithms applied to the accident dataset. Summarizing the prediction efficacy of each method, the GAN-Bayes algorithm proposed in this paper performs well, with a prediction accuracy of 92.46%, and outperforms SVM, RF, GCF and XGBoost-GCF algorithms in all the evaluation metrics. Accordingly, it can be inferred that the GAN-Bayes algorithm performs excellently in predicting the impact of building safety.

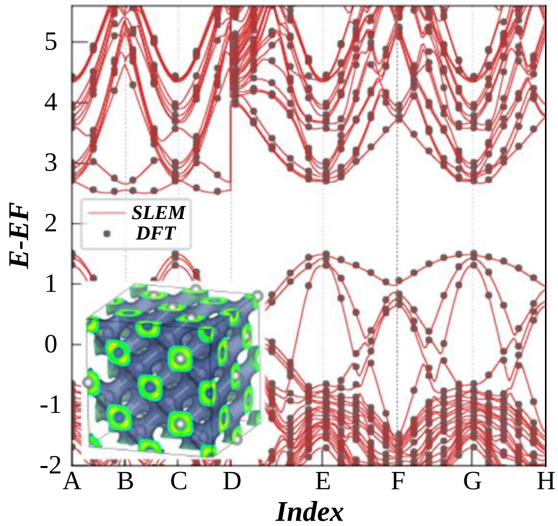

Figure 10: E-EF plots of A-H

The Event - Event Factor graph is an important visualization tool. It aims to visually present the events related to construction safety accidents (E, Event), such as falls from heights, collapses and other accidents themselves, as well as the various factors that induce these events (EF (Event Factor), including personnel illegal operation, equipment aging and failure, harsh environment, etc. By constructing this diagram, the causal correlation and action path between accident events and factors can be clearly sorted out, which can help researchers analyze the accident mechanism more systematically, and then accurately screen the key characteristic variables for the risk prediction model, and improve the accuracy and reliability of the model for the risk prediction of construction engineering safety accidents. Figure 10 reveals the E-EF plot of the A-H algorithm. The overall ROC curve of the GAN-Bayes algorithm and its ROC curves for different impact levels converge to the upper left quadrant, and the corresponding AUC values converge to the ideal value of 1, reflecting the stability and accuracy of the model in predicting the impact levels of various accidents. Compared with the traditional model, the model optimized by feature selection significantly improves accuracy.

To verify GAN-Bayes' ability to deal with incomplete accident features, we randomly deleted some of the leading accident features in the test set to simulate the missing features caused by the untimely data collection during the actual abnormal accidents. The feature missing rate is set to 70%, 50%, 30%, 10%, and 0%, and the corresponding percentage of valid accident features are 30%, 50%, 70%, 90%, and 100%, respectively.

Figure 11 compares the time and memory consumption of different tensor product implementations. It can be seen that even when the feature missing rate is high, the model in this paper still maintains a certain prediction ability. With the increase in the number of effective accident features, the model's prediction performance gradually improves. This result further proves the effectiveness of accident feature enhancement based on GAN, indicating that when an actual accident occurs, with the continuous collection of accident information, the model's prediction accuracy will be continuously improved, providing a more accurate judgment basis for subsequent rescue and traffic diversion.

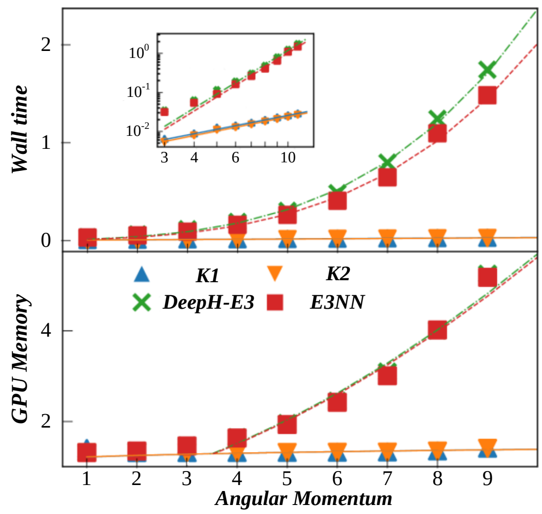

Figure 11: Compares the time and memory consumption of different tensor product implementations

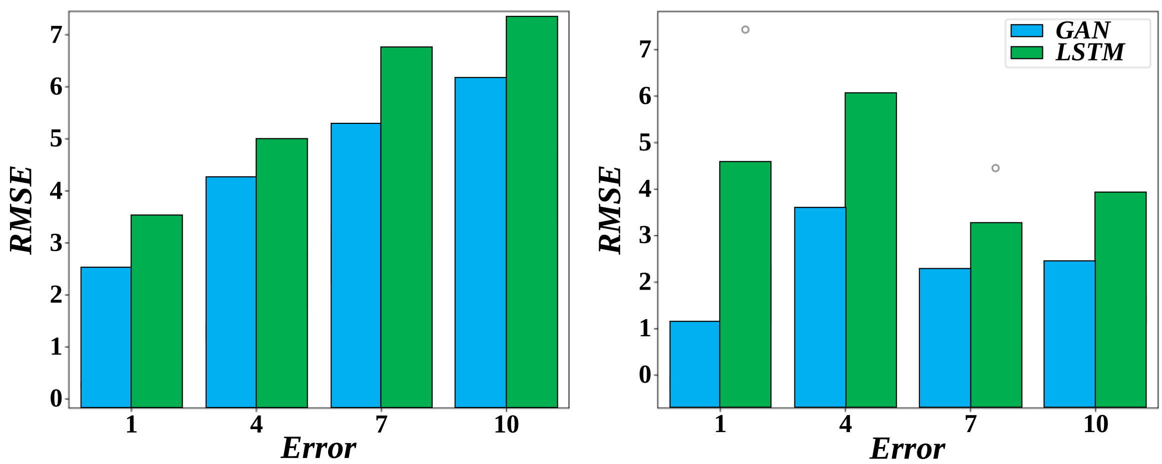

Figure 12: Characteristic importance between accident categories

Figure 12 shows the Root Mean Square Error (RMSE) corresponding to the feature importance between different accident categories in the construction engineering safety accident risk prediction model based on multi-objective optimized GAN network. The blue and green colors used in the figure represent the GAN and LSTM models, respectively, with the horizontal axis representing the Error and the vertical axis representing the RMSE value. As can be seen from the figure, the RMSE values of the two models are different at different error points, reflecting their different performance in dealing with the importance of different features in the risk prediction of construction engineering safety accidents, which is helpful to compare and evaluate the performance of the two models in this study.

7 Conclusion

Through in-depth theoretical discussion and extensive experimental verification, the research on the risk prediction model of construction engineering safety accidents based on multi-objective optimized GAN network has achieved significant research results and important practical application value. This study not only theoretically constructs a multi-objective optimized GAN model that can comprehensively consider multiple dimensions such as data fidelity, model diversity and stability, but also verifies the excellent performance of the model in improving prediction performance through a large number of experiments. This achievement not only provides a more accurate and reliable tool for the risk prediction of construction engineering safety accidents, but also promotes the field to move towards intelligence and science, and contributes an important force to the safety production and sustainable development of the construction industry. The following are the main conclusions of this study:

The prediction model based on multi-objective optimization GAN network proposed in this study shows excellent performance in architectural engineering security incident risk prediction. Compared with traditional prediction methods, the accuracy of this model on the test set has been greatly improved, with an accuracy rate as high as 92.46%. At the same time, it also performs well in key evaluation indicators such as recall rate and F1 value. The results show that the multi-objective optimization GAN network can more effectively capture the complex features of security incident risks and improve the accuracy and reliability of prediction.

By introducing multi-objective optimization strategies, this study not only optimizes the generation and discrimination process of GAN networks, but also makes the model show stronger adaptability and stability in the face of different data sources and complex environments. The improvement of this generalization ability has laid a solid foundation for the wide application of the model in practical engineering projects.

In future research, this study will explore the integration and application prospects of models with the Internet of Things, big data, and other technologies. By building an intelligent and integrated safety accident risk early warning and management system, it provides strong technical support and decision-making basis for safety management in the architectural engineering industry.

References

[1] Mahmoud AlJamal, Ala Mughaid, Bashar Al shboul, Hani Bani-Salameh, Shadi Alzubi, and Laith Abualigah, "Optimizing risk mitigation: A simulation-based model for detecting fake IoT clients in smart city environments," Sustainable Computing: Informatics and Systems, vol. 43, pp. 101019, 2024.

[2] Kemal Haciefendioglu, Fatemeh Mostofi, Vedat Togan, and Hasan Basri Basaga, "CAM-K: a novel framework for automated estimating pixel area using K-Means algorithm integrated with deep learning based-CAM visualization techniques," Neural Computing & Applications, vol. 34, no. 20, pp. 17741-17759, 2022.

[3] Ping Chai, Lei Hou, Guomin Zhang, Quddus Tushar, and Yang Zou, "Generative adversarial networks in construction applications," Automation in Construction, vol. 159, pp. 105265, 2024.

[4] Shanu Verma, Millie Pant, and Vaclav Snasel, "A Comprehensive Review on NSGA-II for Multi-Objective Combinatorial Optimization Problems," Ieee Access, vol. 9, pp. 57757-57791, 2021.

[5] Niloofar Nadim Kabiri, Saeed Emami, and Abdul Sattar Safaei, "Simulation-optimization approach for the multi-objective production and distribution planning problem in the supply chain: using NSGA-II and Monte Carlo simulation," Soft Computing, vol. 26, no. 17, pp. 8661-8687, 2022.

[6] Yali Lv, Jingpu Duan, and Xiong Li, "A survey on modeling for behaviors of complex intelligent systems based on generative adversarial networks," Computer Science Review, vol. 52, pp. 100635, 2024.

[7] Akshay Kumar and T. V. Vijay Kumar, "Multi-Objective Big Data View Materialization Using NSGA-III," International Journal of Decision Support System Technology, vol. 14, no. 1, 2022.

[8] F. Dabbaghi, A. Tanhadoust, M. L. Nehdi, S. Nasrollahpour, M. Dehestani, and H. Yousefpour, "Life cycle assessment multi-objective optimization and deep belief network model for sustainable lightweight aggregate concrete," Journal of Cleaner Production, vol. 318, 2021.

[9] Zixi Hu, Shuang Liu, Fan Yang, Xiaodong Geng, Xiaodi Huo, and Jia Liu, "Research on Multi-objective Optimization Model of Power Storage Materials Based on NSGA-II Algorithm," International Journal of Computational Intelligence Systems, vol. 17, no. 1, 2024.

[10] Tome Sicuaio, Pengxiang Zhao, Petter Pilesjo, Andrey Shindyapin, and Ali Mansourian, "Sustainable and Resilient Land Use Planning: A Multi-Objective Optimization Approach," Isprs International Journal of Geo-Information, vol. 13, no. 3, 2024.

[11] Hongjing Wei et al., "Unified Multi-Objective Genetic Algorithm for Energy Efficient Job Shop Scheduling," Ieee Access, vol. 9, pp. 54542-54557, 2021.
[12] Xuan Xia et al., "GAN-based anomaly detection: A review," Neurocomputing, vol. 493, pp. 497-535, 2022.
[13] Muhammad Ameer and Mohammed Dahane, "NSGA-III-based multi-objective approach for reconfigurable manufacturing system design considering single-spindle and multi-spindle modular reconfigurable machines," International Journal of Advanced Manufacturing Technology, vol. 128, no. 5-6, pp. 2499-2524, 2023.
[14] Zheng Gan, Xiongya Shen, Xiang Liu, and Ying Xie, "Building feature extraction based on natural neighborhood decomposable point feature extraction algorithm," Informatica, vol. 48, no. 22, 2024.
[15] Wejden Gazehi, Rania Loukil, and Mongi Besbes, "Classification of a nanocomposite using a combination between Recurrent Neural Network based on Transformer and Bayesian Network for testing the conductivity property," Expert Systems with Applications, vol. 270, pp. 126518, 2025.
[16] Donghui Shi, Shuling Gan, Jozef Zurada, Jian Guan, Feilong Wang, and Pawel Weichbroth, "A multi-model approach to construction site safety: Fault trees, Bayesian networks, and ontology reasoning," Expert Systems with Applications, vol. 288, pp. 127817, 2025.
[17] Zhongxiang Chang, Zhongbao Zhou, Ruiyang Li, Helu Xiao, and Lining Xing, "Observation scheduling for a state-of-the-art SAREOS: Two adaptive multi-objective evolutionary algorithms," Computers & Industrial Engineering, vol. 169, 2022.
[18] Xiaoyang Liu et al., "Semi-supervised community detection method based on generative adversarial networks," Journal of King Saud University - Computer and Information Sciences, vol. 36, no. 3, pp. 102008, 2024.
[19] Pengda Wang, Zhaowei Liu, Zhanyu Wang, Zongxing Zhao, Dong Yang, and Weiqing Yan, "Graph generative adversarial networks with evolutionary algorithm," Applied Soft Computing, vol. 164, pp. 111981, 2024.
[20] Jian Hu, Yingjun He, Wenqian Xu, Yixin Jiang, Zhihong Liang, and Yiwei Yang, "Anomaly Detection in Network Access-Using LSTM and Encoder-Enhanced Generative Adversarial Networks," Informatica, vol. 49, no. 7, 2025.
[21] Juan Li, "Improved Genetic Algorithm Enhanced with Generative Adversarial Networks for Logistics Distribution Path Optimization," Informatica, vol. 49, no. 11, 2025.
[22] Ima Okon Essiet and Yanxia Sun, "Tracking Variable Fitness Landscape in Dynamic Multi-Objective Optimization Using Adaptive Mutation and Crossover Operators," Ieee Access, vol. 8, pp. 188927-188937, 2020.
[23] Chuyue Zhang and Yuchen Meng, "Bayesian deep learning: An enhanced AI framework for legal reasoning alignment," Computer Law & Security Review, vol. 55, pp. 106073, 2024.
[24] Zhansheng Liu, Chengkuan Ji, Guoliang Shi, and Yanchi Mo, "Structural safety risk prediction method for terminal building steel roof construction considering spatial and temporal variations," Journal of Constructional Steel Research, vol. 224, pp. 109126, 2025.
[25] Jiaqi Wang, Yuqing Fan, Xi Pan, Jun Sun, and Limao Zhang, "Multi-source information fusion for dynamic safety risk prediction of aerial building machine using spatial – temporal multi-graph convolution network," Advanced Engineering Informatics, vol. 65, pp. 103261, 2025.
[26] Claudio Gallicchio and Alessio Micheli, "Architectural richness in deep reservoir computing," Neural Computing & Applications, vol. 35, no. 34, pp. 24525-24542, 2023.
[27] Yair Schwartz, Rokia Raslan, Ivan Korolija, and Dejan Mumovic, "A decision support tool for building design: An integrated generative design, optimisation and life cycle performance approach," International Journal of Architectural Computing, vol. 19, no. 3, pp. 401-430, 2021.
[28] Stephen John Warnett and Uwe Zdun, "Architectural Design Decisions for the Machine Learning Workflow," Computer, vol. 55, no. 3, pp. 40-51, 2022.
[29] Jinmo Rhee, Pedro Veloso, and Ramesh Krishnamurti, "Three decades of machine learning with neural networks in computer-aided architectural design (1990-2021)," Design Science, vol. 9, 2023.
[30] Zhipeng Zhou, Wen Zhuo, Jianqiang Cui, Haiying Luan, Yudi Chen, and Dong Lin, "Developing a deep reinforcement learning model for safety risk prediction at subway construction sites," Reliability Engineering & System Safety, vol. 257, pp. 110885, 2025.

Glossary

Abbreviation	Full Name	Description
TRI	True Risk Influence	Measures the impact of variables on risk levels via the arithmetic mean of High-Risk Influence (HRI) and Low Risk Influence (LRI)
GCF	Generalized Deep Forest	A baseline classification model used for performance comparison with the GAN-Bayes model
XGBoost-GCF	XGBoost Feature Selection-Deep Forest	An integrated model combining XGBoost feature selection with Deep Forest, used in comparative experiments to validate GAN-Bayes effectiveness
FEDPE	Federated Policy Gradient with Byzantine Resilience	A federated learning strategy enhancing model robustness in distributed data scenarios
FEDPG	Federated Policy Gradient	A baseline federated learning strategy for comparing model stability under data missing scenarios
m	Data volume	Used in objective function calculations: V = ∑log D(x), representing the total number of samples
V	Value function	Measures performance of generator/discriminator; core objective function in GAN training
D	Discriminator	Neural network model to distinguish real/generated data; outputs 0 (generated data) or 1 (real data)
G	Generator	Neural network model to produce realistic data; takes random noise Z as input and outputs simulated samples
Z	Random noise input	Input variable for the generator to produce simulated samples
p	Probability distribution	Core GAN objective: making p(G(z)) approximate the real data distribution p_data
Ai/Aj	Attribute variables in GAN-Bayes structure learning	Calculates conditional mutual information I(Ai;Aj|C) to identify variable dependencies
C	Conditional variable/set	Used in conditional probability calculations or as a conditional set in mutual information
A	Attribute variables	Factors such as weather, equipment status as risk factors; measures correlation with risks
TP/TN	True Positive/True Negative samples	Confusion matrix metrics for accuracy calculation
FP/FN	False Positive/False Negative samples	Confusion matrix metrics for recall calculation
α/β	Thresholds for cumulative contribution scores	Thresholds set in the paper to filter key features in accident feature importance analysis

u0	Population mean of differences in paired samples T-test	Null hypothesis parameter in paired T-test
n	Sample size	Number of paired samples or population size in genetic algorithms
d	Difference in paired samples/number of objective functions	Difference in paired samples or number of objective functions for gradient updates
log	Natural logarithm	Used in objective functions and mutual information calculations
E	Complete event set	Used in total probability formula in Bayesian networks