Predicting the Multi-Factor Trigger Probability of Major Urban Security Incidents Using the XGBoost Ensemble Algorithm

Su Feng[1], Guangqing Li[1], Rener Xu[1], Yunlong Zhang[1], Fan Wu[1], Jun Dai[2], Xiang Zhong[2]
[1]Nanjing Emergency Management Bureau, Nanjing 210000, Jiangsu, China
[2]Nanjing Skytech Technology Co., Ltd, Nanjing 210000, Jiangsu, China
Corresponding author: Jun Dai (e-mail: dj991810@126.com)

ABSTRACT Accurate prediction of major urban safety incidents is essential for proactive emergency management and intelligent infrastructure protection, particularly in modern smart cities supported by distributed sensing and communication networks. To address the limitations of conventional statistical and machine learning approaches in modeling nonlinear interactions among heterogeneous risk factors, this study proposes a multi-factor trigger probability prediction framework based on an improved Extreme Gradient Boosting (XGBoost) algorithm. A comprehensive "human– machine–environment–management" factor system is established to characterize complex coupling mechanisms underlying urban safety events, and adaptive sample weighting, feature interaction enhancement, and hierarchical regularization strategies are incorporated to improve prediction performance under imbalanced data conditions. Experimental evaluation using six years of real-world incident data demonstrates that the proposed model achieves an accuracy of 92.3%, a precision of 89.7%, a recall of 91.2%, an F1-score of 90.4%, and an AUC of 0.935, outperforming conventional logistic regression, support vector machine, random forest, and traditional gradient boosting models. Feature importance analysis further identifies extreme weather frequency, equipment aging rate, and personnel violation frequency as the dominant contributors to incident occurrence. Beyond urban emergency management, the proposed framework provides a reliable data-driven methodology for risk assessment and adaptive decision support in electromagnetic sensing networks, wireless monitoring systems, and intelligent communication infrastructures requiring real-time multi-source information fusion.

INDEX TERMS XGBoost, urban safety prediction, multi-factor risk assessment, electromagnetic sensing, wireless monitoring

I. INTRODUCTION

A. RESEARCH BACKGROUND

XGBOOST, an improved gradient-boosted tree model, processes high-dimensional data efficiently, captures nonlinear relationships well, and resists overfitting effectively. It has achieved notable application success in fields such as financial risk prediction and environmental pollutant concentration forecasting. Extending this success, the model holds significant promise for the industry, where it can be leveraged to forecast quality control issues, equipment malfunction risks, and potential safety incidents in complex manufacturing environments. Applying it to predict the multi-factor trigger probability of major urban safety incidents holds promise in overcoming the limitations of traditional models, thereby providing more precise technical support for urban safety management [1-3].

B. RESEARCH SIGNIFICANCE

This study constructs a multidimensional influence factor system encompassing "people-machinery-environment-management," overcoming the limitations of traditional research driven by single or few factors. It enriches the theoretical framework of urban safety incident triggering mechanisms. Simultaneously, by introducing the XGBoost algorithm into incident probability prediction and validating it through comparison with traditional machine learning models, this research clarifies the algorithm's advantages in multi-factor coupling scenarios, offering a new methodological option for safety incident prediction studies [4-6].

The findings can be directly applied to urban safety risk management. Accurate prediction of major incident probabilities helps authorities identify high-risk areas and periods, enabling a shift from passive response to proactive prevention. On the other hand, the model's sensitivity analysis results for key influencing factors can guide management departments in formulating targeted control measures (such as strengthening equipment aging inspections and optimizing personnel training plans), thereby enhancing the efficiency of emergency management resource allocation [7-8]. In the following sections, we define 18 quantifiable impact factors in detail, explain their data sources and numerical methods, and provide a comprehensive basis for high-dimensional feature input in the XGBoost model.

C. CURRENT STATE OF DOMESTIC AND INTERNATIONAL RESEARCH

Regarding predictive models, international research primarily focuses on statistical and machine learning models. Brown et al. employed a logistic regression model to forecast the probability of urban fire incidents, achieving a 78% prediction accuracy rate through factors such as “meteorological conditions + building age + population density”; Garcia et al. developed a comprehensive multi-type safety incident prediction model based on the random forest algorithm, achieving 85% accuracy in an empirical study in Barcelona, Spain. However, the model demonstrated limited predictive capability for low-frequency, high-hazard events (recall rate of only 72%) [9-11]. Therefore, this paper selects logistic regression, SVM, random forest and traditional GBDT as representative baseline models in the experiment, which respectively represent linear separable, maximum interval, nonlinear integration and traditional gradient boosting method. Through fair comparison under the same sample distribution, the advantages of XGBoost in dealing with high-dimensional nonlinear coupling and sample imbalance are empirically tested.

II. MULTI-FACTOR SYSTEM CONSTRUCTION FOR MAJOR URBAN SAFETY INCIDENTS

A. FACTOR SELECTION PRINCIPLES AND METHODS

It should be noted that the ‘major urban security incidents’ mentioned in this article, though low-frequency and high-risk, have traceable causes, observable factors, and clear historical records. Therefore, they fall into the category of ‘gray rhino’ events with statistical predictability, rather than the completely unpredictable ‘black swan’ events as defined by Taleb.

Correlation Principle: Factors must exhibit direct or indirect causal links to triggering major urban safety incidents. Researchers should exclude variables lacking significant association with incident occurrence, using evidence from literature, expert validation, or case verification, or case verification. For instance, indicators like “personnel height” or “equipment color” lack logical relevance to incident triggers and should be excluded outright [12-15].

Quantifiability Principle: Factors must possess clear quantitative standards or data collection channels. Avoid selecting indicators such as “safety culture atmosphere” or “management team cohesion” that are difficult to measure directly and highly subjective. For certain qualitative indicators, quantification must be achieved through reasonable conversion. For instance, “safety system completeness” can be scored quantitatively based on system coverage and clause specificity [16].

Comprehensiveness Principle: The factor system must cover the four core dimensions of “Personnel - Equipment - Environment - Management,” while considering synergistic coupling effects between factors to prevent prediction model bias due to missing dimensions. For instance, focusing solely on “equipment aging” while neglecting “maintenance management” fails to fully capture the evolution of equipment-related risks [17].

Timeliness Principle: Factor data must be obtainable in real-time or near real-time to support dynamic prediction and risk early warning. Data such as extreme weather conditions and equipment operational status require high-frequency updates, while static factors like “city construction age” necessitate adjustments based on annually updated data [18].

B. MULTI-FACTOR SYSTEM STRUCTURE AND QUANTIFICATION STANDARDS

Based on the above screening process, the final multi-factor system for major urban safety incidents still follows the “Human-Machine-Environment-Management” four-dimensional framework [19]. The specific connotations, quantification methods, and data sources for each dimension’s indicators are as follows:

1) Human Dimension

The Human Dimension focuses on “unsafe human behaviors,” encompassing risks stemming from operational errors, insufficient safety awareness, and inadequate skill levels. It comprises four indicators:

Safety Training Coverage Rate: The proportion of personnel in the target area who have completed comprehensive safety training (including pre-job training, periodic refresher training, and specialized training) relative to the total workforce. Quantified as a percentage, calculated as “Number of Trained Personnel / Total Workforce $\times$ 100%.” Data sourced from corporate training records and employee training archives [20].

Personnel Fatigue Level: Measures physiological and psychological fatigue resulting from excessive working hours or irregular schedules. The modified NASA-TLX fatigue assessment scale is used, combined with working hours (daily work >8 hours counts as fatigue accumulation) and night shift frequency (monthly night shifts >4 times counts as high risk) to calculate the fatigue index. The quantification range is 0-1 (0 = no fatigue, 1 = extreme fatigue). Data is sourced from company attendance records and employee health monitoring data [21-22].

2) Machine Dimension

The Machine Dimension focuses on “unsafe machine conditions,” encompassing risks arising from equipment performance degradation, inadequate maintenance, or missing safety devices. It comprises five indicators:

Equipment Aging Rate: The proportion of equipment within the target area exceeding 80% of its design lifespan relative to the total equipment count. For equipment without a defined design lifespan, industry standards are referenced to determine the service life threshold (e.g., 10 years for construction cranes). Quantified as "%", calculated as "Number of Aged Equipment / Total Equipment $\times$ 100%". Data sourced from equipment ledgers and periodic inspection reports [23].

Equipment Failure Frequency: The monthly occurrence of functional failures or abnormal operation events (e.g., shutdown failures, parameter exceedances, component damage). Quantified as "times/month," data sourced from equipment operation logs, fault repair records, and maintenance archives [24].

Equipment Load Factor: The ratio of actual operating load to rated load, reflecting equipment operational intensity. For continuously operating equipment, use the average daily load factor; for intermittently operating equipment, use the peak load factor. Quantified as "%", calculated as "Actual Load / Rated Load $\times$ 100%". Data sourced from equipment monitoring systems and production scheduling records [25].

Infrastructure Aging Level: Specifically refers to the deterioration status of urban public infrastructure (e.g., bridges, pipelines, power lines). Quantified using "Service Life Coefficient $\times$ Damage Severity Coefficient". The service life coefficient is calculated as "Actual Service Life / Design Life," while the damage severity coefficient is determined by on-site inspection results (No Damage = 1.0, Minor Damage = 1.2, Moderate Damage = 1.5, Severe Damage = 2.0). The quantification range is 0-2.0, with data sourced from urban infrastructure census reports and maintenance records [26].

3) Environment Dimension

The Environment Dimension encompasses "external environmental risks," i.e., safety hazards arising from external conditions such as natural and surrounding environments. It comprises four indicators:

Extreme Weather Frequency: Refers to the number of days per month when extreme weather events occur within the target area, including heavy rain (daily rainfall $\geq$50mm), high temperatures (daily maximum temperature $\geq$ 35°C), strong winds (instantaneous wind speed $\geq$10.8m/s), heavy snow (daily snowfall $\geq$10mm), and cold waves (24-hour temperature drop $\geq$ 8°C). Quantified as "days/month," data is sourced from publicly available meteorological records and regional weather observation stations [27].

Peripheral Hazard Density: Refers to the number of hazards within a 1-kilometer radius of the target area, including hazardous chemical warehouses, gas stations, natural gas filling stations, fireworks storage sites, and high-pressure oil pipelines. Quantified as "units/km2," data is sourced from the hazard census database of emergency management departments and Geographic Information System (GIS) spatial analysis results [28].

Emergency Accessway Clearance Rate: Refers to the actual accessibility of emergency accessways (including fire lanes, evacuation routes, and rescue pathways) within the target area. Calculated as "Clear Length / Total Length $\times$ 100%," with clearance standards defined as "no obstructions, no blockages, and widths meeting specifications (fire lanes $\geq$ 4 meters)." Quantified as "%," data sourced from safety inspection records and onsite survey results [29].

4) Management Dimension

The Management Dimension focuses on "the completeness and effectiveness of the safety management system," specifically risks arising from missing management systems, inadequate management measures, or ineffective supervision. It comprises five indicators:

Safety System Completeness: Refers to the coverage ratio of established safety management systems within the target area relative to legal requirements outlined in regulations such as the Work Safety Law and the Urban Safety Risk Control Regulations. Coverage includes eight core modules: responsibility systems, operating procedures, hazard investigation systems, emergency management systems, etc. Each module is scored as "Fully Covered (1 point), Partially Covered (0.5 points), Not Covered (0 points)." The total score divided by 8 multiplied by 100% yields the quantitative result. Quantification is expressed as a percentage, with data sourced from corporate safety management system documents and regulatory filing materials [30].

Emergency Resource Coverage Rate: Refers to the ratio of the coverage area of emergency rescue resources (including firefighting equipment, first-aid devices, emergency supplies, and rescue teams) to the total area of the target zone. Coverage standards include "emergency resource service radius $\leq$ 500 meters" (for firefighting equipment) and "first-aid device response time $\leq$ 5 minutes." Quantified as a percentage, data is sourced from emergency resource census data and GIS spatial analysis results.

Closure Rate of Hazard Rectification: The proportion of identified safety hazards that complete the full closed-loop management cycle of "identification-registration-rectification-inspection-closure." Rectification deadlines are determined by hazard severity (general hazards $\leq$7 days, major hazards $\leq$24 hours). Quantified as a percentage, calculated as "Number of hazards with closed-loop rectification / Total number of identified hazards $\times$ 100%." Data sourced from hazard investigation and management ledgers.

Safety Investment Intensity: The annual expenditure on safety management within the target area as a percentage of the annual total output value or total budget, including expenditures on safety facility procurement, training and education, hazard rectification, emergency drills, insurance costs, etc. Quantified as a percentage (%), data sourced from corporate financial statements and specialized safety investment ledgers.

C. FACTOR IMPORTANCE RANKING AND CORE FACTOR IDENTIFICATION

To further focus on key influencing factors and reduce model computational complexity, the 18 factors require importance ranking to identify core factors. This study employs a combined approach of improved Grey Relational Analysis (GRA) and Random Forest feature importance assessment to achieve quantitative factor ranking.

1) Improved Grey Relational Analysis (GRA)

Gray Relational Analysis assesses a factor's influence on the target variable by calculating its correlation with a reference sequence. Traditional GRA suffers from subjective determination of the discrimination coefficient when handling multidimensional data. This paper introduces the entropy weight method to determine the discrimination coefficient ζ, enhancing the objectivity of analysis results:

Data Standardization: To eliminate the impact of different measurement units, Min-Max normalization is applied to convert raw data to the [0,1] range. For positive factors (e.g., safety training coverage, where higher values indicate lower risk), use formula (1); for negative factors (e.g., frequency of non-compliant operations, where higher values indicate higher risk), use formula (2):

$$x'_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)} \tag{1}$$

$$x'_{ij} = \frac{\max(x_j) - x_{ij}}{\max(x_j) - \min(x_j)} \tag{2}$$

Where x'_{ij} is the normalized value of the jth factor for the ith sample, x_{ij} is the raw data of the jth factor, and $\min(x_j)$ and $\max(x_j)$ are the minimum and maximum values of the jth factor, respectively.

Define the reference sequence: Use "whether a major urban security incident occurred" as the reference sequence Y, where occurrence is denoted as 1 and non-occurrence as 0, i.e., $Y = [y_1, y_2, \ldots, y_n]$ (n is the number of samples).

Calculate the entropy-weighted discrimination coefficient ζ: First compute the information entropy H_j for each factor using formula (3); then calculate the weights w_j based on the information entropy using formula (4); finally determine $\zeta = 1 - \max(w_j)$ to preserve the influence of factors with the largest weights on the association calculation:

$$H_j = -\frac{1}{\ln n} \sum_{i=1}^{n} x'_{ij} \ln x'_{ij} \tag{3}$$

$$w_j = \frac{(1 - H_j)}{\sum_{j=1}^{18} (1 - H_j)} \tag{4}$$

Calculate the grey correlation degree r_j: The formula is (5). A higher correlation degree indicates a stronger association between the factor and the event trigger:

$$r_j = \frac{\sum_{i=1}^{n} |y_i - x'_{ij}| + \zeta \max_j \max_i \sum_i |y_i - x'_{ij}|}{\min_j \min_i \sum_i |y_i - x'_{ij}| + \zeta \max_j \max_i \sum_i |y_i - x'_{ij}|} \tag{5}$$

III. CONSTRUCTION OF A MULTI-FACTOR TRIGGER PROBABILITY PREDICTION MODEL BASED ON XGBOOST

A. XGBOOST ALGORITHM PRINCIPLES AND IMPROVEMENTS

1) Fundamental Principles

XGBoost (Extreme Gradient Boosting), proposed by Chen & Guestrin in 2016, is an ensemble learning algorithm optimized from Gradient Boosting Decision Trees (GBDT). Its core advantage lies in progressively correcting prediction errors through the serial integration of multiple decision trees, while incorporating regularization mechanisms and a parallel computing framework to balance prediction accuracy and training efficiency. Its core concept is: based on an "additive model," each iteration constructs a new decision tree to maximize gradient descent on the overall loss function. The final prediction result is obtained by the weighted sum of all decision trees.

XGBoost's objective function comprises a loss function and a regularization term, as shown in Formula (6):

Obj(θ) = Σ_(i=1)^n l(y_i, y^i(t)) + Σ_(k=1)^k Ω(f_k) (6)

Here, $Obj(\theta)$ represents the objective function for it-eration t, l(yi,y^i(t)) is the loss function measuring the deviation between the true value y_i and the prediction $\hat{y}_i^{(t)}$ at iteration t; $\Omega(f_k)$ is the regularization term controlling the complexity of the kth decision tree to prevent overfitting; θ is the set of model parameters encompassing the structure and weight parameters of all decision trees.

For binary classification problems such as predicting the probability of triggering major urban security incidents, this paper employs the logistic loss function as the objective function, as shown in Formula (7):

l(yi,y^i(t))=−[yilnσ(y^i(t))+(1−yi)ln(1−σ(y^i(t)))] (7)

Here, $\sigma(\cdot)$ denotes the Sigmoid activation function, which maps the prediction value to the interval [0,1] as the event occurrence probability output, as shown in Formula (8):

σ(y^i(t)) = 1 + e^(−y^i(t)) / 1 (8)

XGBoost employs a greedy algorithm to construct decision trees. At each split, it selects the optimal splitting feature and split point by calculating the "gain value," defined by Formula (9):

$$Gain = \frac{1}{2}\left[\frac{G_L^2}{H_L+\lambda} + \frac{G_R^2}{H_R+\lambda} - \frac{(G_L+G_R)^2}{H_L+H_R+\lambda}\right] - \gamma \quad (9)$$

Here, G_L and H_L represent the gradient sum and Hessian sum of samples in the left subtree, respectively, while G_R and H_R denote the gradient sum and Hessian sum of samples in the right subtree. Splitting

proceeds if Gain > 0, reducing the objective function value; otherwise, splitting halts.

2) Algorithm Improvement Strategy

Addressing the characteristics of multi-factor coupling in major urban security incidents and imbalanced sample distribution (low proportion of positive samples), this paper enhances the traditional XGBoost algorithm in three aspects—sample balancing, feature optimization, and model regularization—to improve the model's generalization capability and prediction accuracy:

Sample Weighting Balancing Strategy: Due to the low frequency of major safety incidents, positive samples (event triggers) constitute only 5.8% (1263/21600) of the dataset. Traditional models tend to favor predicting negative samples, resulting in low recall rates for positive samples. This paper employs a "category weight adaptive adjustment" method to dynamically set category weights based on sample distribution: Set the negative sample weight to 1 and the positive sample weight to N_{neg}/N_{pos} (where N_{neg} is the number of negative samples and N_{pos} is the number of positive samples). By introducing the sample weight factor w_i into the loss function, as shown in Formula (10), the model's focus on positive samples is enhanced:

l(yi,y^i(t))=−wi[yilnσ(y^i(t))+(1−yi)ln(1−σ(y^i(t)))] (10)

Feature Interaction Enhancement Module: Major urban safety incidents are often triggered by the coupling of multiple factors (e.g., "extreme weather + equipment aging," "human violations + management loopholes"). Traditional XGBoost can only indirectly capture feature interactions through decision tree splits, yielding limited effectiveness. This paper introduces a "feature cross-term construction" mechanism. By expanding polynomial features, it generates second-order interaction features (e.g., "extreme weather frequency × equipment aging rate," "personnel violation frequency × safety system completeness"), adding 28 new interaction features. This enriches model input information and enhances the ability to characterize multi-factor coupling effects.

Hierarchical Regularization Optimization: Traditional XGBoost employs uniform regularization parameters γ and λ, which struggle to adapt to varying feature dimensionality complexities. This paper adopts a "hierarchical regularization" strategy: setting distinct regularization parameters for raw features (γ_1, λ_1) and interaction features (γ_2, λ_2). Grid search optimizes parameter combinations to prevent overfitting in raw features while avoiding redundant information introduc-tion in interactions.

B. MODEL CONSTRUCTION PROCESS

Based on the improved XGBoost algorithm, this paper constructs a multi-factor trigger probability prediction model for major urban security incidents. The specific process comprises four stages: data preprocessing, model training, hyperparameter optimization, and model validation.

Core Feature Selection: Based on the 16 core factors identified in Chapter 2, combined with 28 second-order interaction features generated by the feature interaction enhancement module, an initial feature set comprising 44 features is formed. The input features of the model all meet the principle of 'measurable before the event occurs', among which 12 are daily or real-time dynamic data, and the rest are monthly update indicators.

Feature Importance Ranking: Utilizing the improved XGBoost built-in feature importance evaluation function (based on cumulative split gain values), importance scores are calculated for all 44 features and ranked from highest to lowest.

Redundant Feature Elimination: Set a feature importance threshold of 0.01 (i.e., features scoring below 1% of the total score were deemed redundant). Eight low-importance interaction features (e.g., "Safety Assessment Pass Rate × Emergency Resource Coverage Rate," "Equipment Load Rate × Environmental Temperature/Humidity Deviation") were removed, ultimately retaining 36 features for model training. This approach ensures feature information integrity while reducing computational complexity.

Dataset Partitioning: Employed a "stratified sampling" method to divide the preprocessed dataset into training, validation, and test sets at a 7:2:1 ratio. This ensured consistent positive sample proportions (5.8% each) across all three datasets, preventing sample distribution bias from affecting model evaluation. After partitioning, the training set contained 15,120 samples (884 positive, 14,236 negative), the validation set contained 4,320 samples (253 positive, 4,067 negative), and the test set contained 2,160 samples (126 positive, 2,034 negative).

Hyperparameter Optimization: Hyperparameters significantly impact the predictive performance of XGBoost models. This study selected key hyperparameters (Table 1) and employed a "grid search + 5-fold cross-validation" approach for optimization. The F1 score (a composite measure of precision and recall) on the validation set served as the optimization objective to identify the optimal hyperparameter combination.

TABLE 1. Key Hyperparameters and Optimization Ranges for XGBoost Models

Hyperparameter Name	Meaning	Optimization Range
learning_rate	Learning rate	[0.01, 0.05, 0.1, 0.2]
max_depth	Maximum depth of decision tree	[3, 5, 7, 9, 11]
n_estimators	Number of decision trees	[100, 200, 300, 400, 500]
subsample	Subsample ratio of training data	[0.7, 0.8, 0.9, 1.0]
colsample_bytree	Subsample ratio of features	[0.7, 0.8, 0.9, 1.0]
gamma	Complexity penalty coefficient of tree	[0, 0.1, 0.2, 0.5, 1.0]
lambda	L2 regularization coefficient	[0.1, 1, 5, 10, 20]
alpha	L1 regularization coefficient	[0, 0.1, 1, 5, 10]

The optimal hyperparameter combination determined through grid search and cross-validation is shown in Table 2:

TABLE 2. Optimal Hyperparameter Combination

Hyperparameter Name	Optimal Value
learning_rate	0.08
max_depth	7
n_estimators	300
subsample	0.8
colsample_bytree	0.9
gamma	0.3
lambda	5
alpha	1

Model Training: An improved XGBoost model was trained on the training set using the optimal hyperparameter combination. The training process employed an "early stopping mechanism": training automatically halted when the validation set's F1 score failed to improve for 10 consecutive iterations, preventing overfitting caused by overtraining. The final model training reached 246 iterations, with a training duration of 18.3 minutes (on an Intel Core i7-12700H processor with 16GB RAM). The final classification threshold was determined by maximizing the F1 score on the validation set PR curve, set at 0.38 to balance high recall and reasonable precision.

To comprehensively evaluate model performance, this paper selected four commonly used classification model evaluation metrics and conducted quantitative analysis in conjunction with the confusion matrix (Table 3):

TABLE 3. Confusion Matrix for Binary Classification Problems

Actual / Predicted	Positive (Event Triggered)	Negative (Event Not Triggered)
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

Accuracy: The proportion of correctly predicted samples to the total number of samples, reflecting the model's overall predictive capability. The formula is shown in (11):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

Precision: The proportion of true positive samples among those predicted as positive, reflecting the reliability of the model's predictions. The formula is shown in (12):

$$Precision = \frac{TP}{TP + FP} \tag{12}$$

Recall: The proportion of true positive samples correctly predicted as positive, reflecting the model's ability to identify event triggers (avoiding false negatives). The formula is shown in (13):

$$Recall = \frac{TP}{TP + FN} \tag{13}$$

F1 Score: The harmonic mean of precision and recall, providing a comprehensive measure of the model's balanced performance. The formula is shown in (14):

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{14}$$

Additionally, the ROC curve (Receiver Operating Characteristic curve) and AUC value (Area Under the ROC Curve) are used to evaluate the model's discrimination capability. An AUC value closer to 1 indicates better classification performance.

IV. EMPIRICAL ANALYSIS AND RESULTS DISCUSSION

A. DATA SOURCES AND PREPROCESSING RESULTS

All data used in this empirical analysis were obtained from public channels and authoritative databases to ensure authenticity and reliability:

1) Event Data

- Ministry of Emergency Management's "National Direct Reporting System for Production Safety Accident Statistics" Major safety incidents recorded from 2018 to 2023, filtered based on the criteria of "causing casualties exceeding 10 persons or direct economic losses exceeding 5 million yuan." The dataset comprises

1,263 valid incident records across 30 cities, including: construction collapses (245, 19.4%), bridge/road collapses (156, 12.3%), and special equipment explosions (154, 12.2%). The balanced sample distribution aligns with the actual occurrence patterns of major urban safety incidents.

B. MODEL PERFORMANCE EVALUATION RESULTS

1) Overall Performance

The performance metrics of the improved XGBoost model and four comparison models on the test set are shown in Table 4. All models were evaluated under identical sample distributions, with XGBoost employing a positive sample weighting strategy to mitigate imbalanced data.

Table 4 Comparison of Test Set Performance Metrics Across Models The improved XGBoost model significantly outperforms all baseline models across every metric (Table 4): its accuracy reaches 92.3%, surpassing logistic regression by 9.7 percentage points and traditional GBDT by 2.2 percentage points; Precision and recall are 89.7% and 91.2%, respectively, ensuring prediction reliability while effectively reducing the risk of missed events; the F1 score reaches 90.4%, demonstrating outstanding overall performance; the AUC value is 0.935, approaching the ideal value of 1, indicating the model's exceptional ability to distinguish between positive and negative samples.

TABLE 4. Comparison of Performance Metrics Across Model Test Sets

Model	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression (LR)	82.6%	75.3%	71.4%	73.3%	0.786
Support Vector Machine (SVM)	85.7%	78.9%	76.2%	77.5%	0.823
Random Forest (RF)	88.9%	83.5%	81.7%	82.6%	0.867
Traditional GBDT	90.1%	85.2%	84.9%	85.0%	0.892
Improved XGBoost (This Study)	92.3%	89.7%	91.2%	90.4%	0.935

2) Predictive Performance Across Event Types

To validate the model's generalization capability, the test set was segmented by event type to evaluate the improved XGBoost model's predictive effectiveness for different categories of major security incidents. Results are presented in Table 5:

TABLE 5. Improving the predictive performance of XGBoost models for different event types

Event Type	Accuracy	Recall	F1-Score
Large-Scale Fire	93.1%	92.5%	92.8%
Hazardous Chemical Leak	91.8%	90.7%	91.2%
Construction Collapse	92.5%	91.4%	91.9%
Bridge and Road Collapse	90.7%	88.5%	89.6%
Special Equipment Explosion	91.3%	89.8%	90.5%

The results indicate that the model achieves prediction accuracy exceeding 90% for all event categories, with the highest performance observed for large-scale fires (F1 score: 92.8%). Prediction accuracy for bridge and road collapses is relatively lower (F1 score: 89.6%), though the overall variance remains minimal. This demonstrates the model's effective adaptation to the multi-factor triggering characteristics of various major urban safety incidents, reflecting strong generalization capabilities.

C. KEY RISK FACTOR ANALYSIS BASED ON XGBOOST FEATURE IMPORTANCE

This study employs XGBoost's built-in gain-based global feature importance assessment method, which effectively reflects the average contribution of each factor to prediction outcomes, providing actionable insights for management interventions. By enhancing the feature importance evaluation function of the XGBoost model, we obtained a ranked list of 16 core factors (Table 6), identifying the key elements influencing the triggering of major urban safety incidents:

TABLE 6. Top 10 Feature Importance Ranking Based on XGBoost Gain

Rank	Factor Name	Importance Score	Dimension	Rank
1	Frequency of Extreme Weather	0.186	Environment	1
2	Equipment Aging Rate	0.153	Machine	2
3	Frequency of Personnel Violations	0.147	Human	3
4	Completeness of Safety System	0.112	Management	4
5	Frequency of Equipment Failures	0.098	Machine	5
6	Closed-Loop Rate of Hidden Danger Rectification	0.085	Management	6
7	Coverage Rate of Safety Training	0.076	Human	7
8	Density of Surrounding Hazards	0.063	Environment	8
9	Coverage Rate of Emergency Resources	0.052	Management	9
10	Equipment Load Rate	0.048	Machine	10

The factor importance analysis results indicate:

The environmental dimension's "Extreme Weather Frequency" (0.186), the equipment dimension's "Equipment Aging Rate" (0.153), and the personnel dimension's "Frequency of Personnel Violations" (0.147) are the three core factors influencing incident triggering. Their cumulative importance accounts for 48.6%, indicating that natural environmental risks, equipment operational status, and personnel operational behavior are key levers for preventing and controlling major urban safety risks.

Three factors from the equipment dimension entered the Top 10, with a cumulative importance score of 0.299. This indicates that equipment safety status has the most concentrated impact on incident triggering, necessitating enhanced equipment aging detection and fault early warning systems.

The management dimension factors "Safety System Completeness" (0.112) and "Closure Rate of Hidden Hazard Rectification" (0.085) ranked highly, validating the crucial role of a robust safety management system in risk prevention and control.

V. CONCLUSION

Addressing the challenges of multi-factor coupling and insufficient accuracy in traditional models for predicting major urban safety incidents, this study constructs a 16-factor core system across four dimensions —'Person-Machine-Environment-Management'—and proposes a multi-factor trigger probability prediction model based on an improved XGBoost algorithm. We specifically tailor this robust framework to the industry, enhancing the prediction accuracy for complex hazards like machinery-related accidents and chemical storage risks within manufacturing facilities. The model's effectiveness is validated using six years of empirical data from 30 cities. Research findings indicate: The improved XGBoost model achieved accuracy, precision, recall, and F1 scores of 92.3%, 89.7%, 91.2%, and 90.4% respectively, outperforming traditional models like logistic regression and SVM across all metrics. With an AUC value of 0.935, it demonstrates strong classification capabilities. Extreme weather frequency, equipment aging rate, and personnel violation frequency emerged as the three core factors influencing incident triggering, collectively accounting for 48.6% of cumulative importance. The multi-factor framework comprehensively covers key event-triggering factors. The enhanced XGBoost model effectively addresses challenges of sample imbalance and multi-factor coupling modeling. This research provides crucial scientific support for the tiered management of urban safety risks and the pre-deployment of emergency resources across urban systems. Significantly, when applied to a high-risk manufacturing environment such as the industry, this methodology becomes instrumental. Specifically, it facilitates the development of proactive safety protocols for manufacturing complexes by offering a new methodological reference for multi-factor-driven safety event prediction research.

AUTHOR CONTRIBUTIONS

Conceptualization –Su Feng, Guangqing Li, Rener Xu, Yunlong Zhang, Fan Wu, Jun Dai and Xiang Zhong; methodology – Su Feng, Guangqing Li, Rener Xu, Fan Wu, Jun Dai and Xiang Zhong; investigation – Su Feng, Guangqing Li, Rener Xu, Yunlong Zhang, Fan Wu, Jun Dai and Xiang Zhong; writing-original draft preparation – Su Feng, Guangqing Li, Rener Xu, Yunlong Zhang, Fan Wu, Jun Dai and Xiang Zhong. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] Y. Zhao, F. M. Ramos, and B. Li, "Retraction Note: Integrated framework to integrate spark-based big data analytics and for health monitoring and recommendation in sports using XGBoost algorithm," Soft Computing, vol. (prepublish), pp. 1-1, 2026, doi: 10.1007/s00500-026-11199-w.

[2] E. Dastres and M. Edalat, "Deep ensemble learning for weed risk mapping: Hybrid RF-CatBoost and CNN-XGBoost algorithms for predicting Chenopodium album distribution in rapeseed fields," Smart Agricultural Technology, vol. 13, Art. no. 101731, 2026, doi: 10.1016/j.atech.2025.101731.

[3] R. Suppakul, W. Jitchaijaroen, S. Keawsawasvong, et al., "Undrained uplift capacity prediction of open-caisson anchors in anisotropic clays using XGBoost integrated with mutation-based genetic algorithms," Artificial Intelligence in Geosciences, vol. 6, no. 2, Art. no. 100162, 2025, doi: 10.1016/j.aiig.2025.100162.

[4] F. Khodadadi, F. Taghizadeh, H. A. Baghi, et al., "Leveraging ensemble machine learning models (XGBoost and random forest) and genetic algorithms to predict factors contributing to the liposomal entrapment of therapeutics," Nanoscale, 2025, doi: 10.1039/D5NR01741F 3753.

[5] F. D. Muhyi and O. Ata, "Integrative machine learning approaches for enhanced cardiovascular disease prediction: a comparative analysis of XGBoost and ANFIS algorithms," The Journal of Supercomputing, vol. 81, no. 12, pp. 1213, 2025, doi: 10.1007/s11227-025-07687-9.

[6] Y. Chen, Y. Zhang, C. Li, et al., "Application of XGBoost Model Optimized by Multi-Algorithm Ensemble in Predicting FRP-Concrete Interfacial Bond Strength," Materials, vol. 18, no. 12, pp. 2868, 2025, doi: 10.3390/ma18122868.

[7] H. Huisi, Y. Yiming, H. Jianlong, et al., "Construction of a Clinical Path Discrimination Model for Stroke Patients Based on the XGBoost Integrated Learning Algorithm and Its Application Analysis in the MOP under the DIP Payment Model," Journal of Clinical and Nursing Research, vol. 9, no. 4, pp. 291-298, 2025, doi: 10.26689/jcnr.v9i4.10484.

[8] H. Ahmad, Z. Yinghua, M. Khan, et al., "Morphometric assessment and soil erosion susceptibility maping using ensemble extreme gradient boosting (XGBoost) algorithm: a study for Hunza-Nagar catchment, Northern Pakistan," Environmental Earth Sciences, vol. 83, no. 21, pp. 605, 2024, doi: 10.1007/s12665-024-11909-3.

[9] Z. Zou, B. Wang, X. Hu, et al., "Enhancing requirements-to-code traceability with GA-XWCoDe: Integrating XGBoost, Node2Vec, and genetic algorithms for improving model performance and stability," Journal of King Saud University - Computer and Information Sciences, vol. 36, no. 8, Art. no. 102197, 2024, doi: 10.1016/j.jksuci.2024.102197.

[10] S. Luo, B. Wang, Q. Gao, et al., "Stacking integration algorithm based on CNN-BiLSTM-Attention with XGBoost for shortterm electricity load forecasting," Energy Reports, vol. 12, pp. 2676-2689, 2024, doi: 10.1016/j.egyr.2024.08.078.

[11] W. Xia, B. Liu, and H. Xiang, "Prediction of Liquid Accumulation Height in Gas Well Tubing Using Integration of Crayfish Optimization Algorithm and XGBoost," Processes, vol. 12, no. 9, pp. 1788, 2024, doi: 10.3390/pr12091788.

[12] W. Li, Y. Peng, and K. Peng, "Diabetes prediction model based on GA-XGBoost and stacking ensemble algorithm," PloS one, vol. 19, no. 9, Art. no. e0311222, 2024, doi: 10.1371/journal.pone.0311222.

[13] A. Cao, X. Jin, Y. Wang, et al., "The effects of suspension-supported training on dynamic balance capacity in stroke patients: a systematic review and meta-analysis enhanced by XGBoost machine learning," Frontiers in Medicine, vol. 13, Art. no. 1747067, 2026, doi: 10.3389/fmed.2026.1747067.

[14] A. Saghir, R. Basit, H. Lal, et al., "The Deep Learning ResNet101 and Ensemble XGBoost Algorithm with Hyper parameters Optimization Accurately Predict the Lung Cancer," Applied Artificial Intelligence, pp. 37(1), 2023, doi: 10.1080/08839514.2023.2166222.

[15] Z. W. Hua, Y. L. Fei, W. Tao, et al., "Intelligent prediction model of mechanical properties of ultrathin niobium strips based on XGBoost ensemble learning algorithm," Computational Materials Science, pp. 231, 2024, doi: 10.1016/ j.commatsci.2023.112579 3754.

[16] J. O. Abiodun and I. A. Wreford, "Stroke Prediction Using Smote for Data Balancing, XGBoost and KNN Ensemble Algorithms," Journal of Applied Physical Science International, pp. 42-53, 2023, doi: 10.56557/japsi/2023/v15i18349.

[17] K. Dheeraj, S. S. Kumar, and K. R. Singh, "Early health prediction framework using XGBoost ensemble algorithm in intelligent environment," Artificial Intelligence Review, vol. 56, no. Suppl 1, pp. 1591-1615, 2023, doi: 10.1007/s10462-023-10565-6.

[18] E. Kim and I. Joe, "Handover Triggering Prediction with the Two-Step XGBOOST Ensemble Algorithm for Conditional Handover in Non-Terrestrial Networks," Electronics, pp. 12(16), 2023, doi: 10.3390/electronics12163435.

[19] D. Selçuk and E. S. Kutlug, "Predicting occurrence of liquefaction-induced lateral spreading using gradient boosting algorithms integrated with particle swarm optimization: PSO-XGBoost, PSO-LightGBM, and PSO-CatBoost," Acta Geotechnica, vol. 18, no. 6, pp. 3403-3419, 2023, doi: 10.1007/s11440-022-01777-1.

[20] D. Selçuk and E. S. Kutlug, "An investigation of feature selection methods for soil liquefaction prediction based on treebased ensemble algorithms using AdaBoost, gradient boosting, and XGBoost," Neural Computing and Applications, vol. 35, no. 4, pp. 3173-3190, 2022, doi: 10.1007/s00521-022-07856-4.

[21] M. Zivkovic, N. Bacanin, M. Antonijevic, et al., "Correction: Zivkovic et al," Hybrid CNN and XGBoost Model Tuned by Modified Arithmetic Optimization Algorithm for COVID-19 Early Diagnostics from X-Ray Images. Electronics 2022, 11, 3798. Electronics, vol. 15, no. 4, pp. 727, 2026, doi: 10.3390/electronics15040727.

[22] C. Yang, J. Li, K. Zhou, et al., "CCO-XGBoost Hybrid Model for Prediction of Blasting-Induced Peak Particle Velocity in Open-Pit Mines: A SHAP-Driven Sensitivity Analysis," Mathematics, vol. 14, no. 4, pp. 596, 2026, doi: 10.3390/math14040596.

[23] L. C. Xu, B. Yuan, X. Hu, et al., “Estimation of Hourly PM2.5 Mass Concentration in Guanzhong Based on Spatio-temporal XGBoost Model,” Huan jing ke xue= Huanjing kexue, vol. 47, no. 2, pp. 663-672, 2026, doi: 10.13227/J.HJKX.202412262.

[24] N. H. Bui, S. Keawsawasvong, N. H. Le, et al., “Hybrid XGBoost - 3D FEA approach for predicting bearing capacity of rectan gular foundations on rock slopes,” Rock Mechanics Bulletin, vol. 5, no. 3, Art. no. 100274, 2026, doi: 10.1016/j.rockmb.2025.100274.

[25] J. Noh and S. S. Kang, “Evaluation of prediction models for slope stability on road cut slopes using XGBoost, LightGBM, and CatBoost,” KSCE Journal of Civil Engineering, vol. 30, no. 3, Art. no. 100379, 2026, doi: 10.1016/j.kscej.2025.100379.

[26] J. Zhao, X. Du, J. Lin, et al., “Determinants and prediction of home nursing utilization among older adults in China: an integration of logistic regression and XGBoost algorithm,” BMC nursing, 2026, doi: 10.1186/s12912-026-04371-y.

[27] C. Hazman, A. Guezzaz, S. Benkirance, et al., “A smart model integrating LSTM and XGBoost for improving LoT-enabled smart cities security,” Cluster Computing, vol. 28, no. 1, pp. 70, 2024, doi: 10.1007/s10586-024-04780-1.

[28] K. Shubham, S. Metya, K. A. Sinha, et al., “XGBoost based surrogate technique for system reliability analysis of foundation over cavity aided with bootstrapping,” Scientific reports, vol. 16, no. 1, pp. 7113, 2026, doi: 10.1038/s41598-026-37058-0.

[29] J. Zeng, Y. Zhu, F. Ye, et al., “Application of XGBoost and logistic regression in predicting 90 days mortality for elderly severe acute renal failure patients,” Scientific reports, vol. 16, no. 1, pp. 7077, 2026, doi: 10.1038/s41598-026-37828-w.

[30] L. Kuang, Y. Zeng, C. Hu, et al., “Fault Diagnosis of Hydro-Power Units Using BP Neural Network and XGBoost Algorithm for Enhanced Operational Safety,” Processes, vol. 14, no. 3, pp. 517, 2026, doi: 10.3390/pr14030517 3756.