applied sciences

MDPI

Article

Dynamic Fire Risk Classification Prediction of Stadiums: Multi-Dimensional Machine Learning Analysis Based on Intelligent Perception

Ying Lu [1,2,*], Xiaopeng Fan [1], Zhipan Zhao [1] and Xuepeng Jiang [1]

[1] School of Resource and Environmental Engineering, Wuhan University of Science and Technology, Wuhan 430081, China; fan_xiaopeng@163.com (X.F.); zhao_zhipan@163.com (Z.Z.); jxp5276@126.com (X.J.)
[2] Hubei Industrial Safety Engineering Technology Research Center, Wuhan 430081, China
* Correspondence: luying@wust.edu.cn

Abstract: Stadium fires can easily cause massive casualties and property damage. The early risk prediction of stadiums will be able to reduce the incidence of fires by making corresponding fire safety management and decision making in an early and targeted manner. In the field of building fires, some studies apply data mining techniques and machine learning algorithms to the collected risk hazard data for fire risk prediction. However, most of these studies use all attributes in the dataset, which may degrade the performance of predictive models due to data redundancy. Furthermore, machine learning algorithms are numerous and applied to fewer stadium fires, and it is crucial to explore models suitable for predicting stadium fire risk. The purpose of this study was to identify salient features to build a model for predicting stadium fire risk predictions. In this study, we designed an index attribute threshold interval to classify and quantify different fire risk data. We then used Gradient Boosting-Recursive Feature Elimination (GB-RFE) and Pearson correlation analysis to perform efficient feature selection on risk feature attributes to find the most informative salient feature subsets. Two cross-validation strategies were employed to address the dataset imbalance problem. Using the smart stadium fire risk data set provided by the Wuhan Emergency Rescue Detachment, the optimal prediction model was obtained based on the identified significant features and six machine learning methods of 12 combination forms, and full features were input as an experimental comparison study. Five performance evaluation metrics were used to evaluate and compare the combined models. Results show that the best performing model had an F1 score of 81.9% and an accuracy of 93.2%. Meanwhile, by introducing a precision-recall curve to explain the actual classification performance of each model, AdaBoost achieves the highest Auprc score (0.78), followed by SVM (0.77), which reveals more stable performance under such imbalanced data.

Keywords: machine learning; dynamic fire risk; stadium; Internet of Things (IoT); data mining

check for updates

Citation: Lu, Y.; Fan, X.; Zhao, Z.; Jiang, X. Dynamic Fire Risk Classification Prediction of Stadiums: Multi-Dimensional Machine Learning Analysis Based on Intelligent Perception. *Appl. Sci.* **2022**, *12*, 6607. https://doi.org/10.3390/app12136607

Academic Editor: Federico Divina

Received: 8 June 2022
Accepted: 27 June 2022
Published: 29 June 2022

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

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

1. Introduction

As the carrier of various cultural and recreational activities, stadiums have many internal facilities, a large flow of people, closed spaces, complex structures, uneven personnel quality, weak awareness of fire prevention and disaster prevention, hidden safety hazards always exist, and a high fire risk. In recent years, the stadiums in China have suffered more than 200 fires, with the deaths of more than 1000 people, which fully reveals the serious problem in the security management of stadiums [1].There are many risk factors for fires in stadiums; the mechanism is complex, and there may be a linear correlation between internal factors, resulting in inaccurate assessment results. In response to this deficiency, existing research also attempts to use some methods to overcome the high-dimensional problem (the curse of dimensionality). Hamed et al. [2] proposed a feature selection method based on recursive feature addition and gram techniques and tested it on the ISCX2012

dataset, and the results showed that the performance of the model was significantly improved. Latah et al. [3] used principal component analysis for dimensionality reduction and evaluated their models, which outperformed traditional supervised machine learning (ML) algorithms in terms of accuracy, false positives, and recall.

At present, static assessment is the main method for the prediction of fire risk in stadiums, and mathematical statistical methods such as the fuzzy evaluation method [4], Analytic Hierarchy Process (AHP) [5], and structural entropy weight method [6] are mostly used. The selection of the indicator weight is subjective, and some indicators require on-site scoring by experts, which leads to the inability to verify the accuracy of the static evaluation model. Choi et al. [5] ranked and classified various fire factors in urban residential areas based on AHP, and designed a new tool for residential fire risk probability. They developed a fire risk prediction model and a GIS risk hazard map with fire factor classification settings. Liu et al. [6] introduced the structural entropy weight method into the index weight determination process, and established a new large-scale commercial building fire risk assessment system, which can be used as the input data feature to predict the fire safety performance of the building. However, none of them circumvent the subjectivity of indicator quantification and weight assignment.

Some scholars have paid more attention to the advantages of ML algorithms in solving the subjectivity of fire risk assessment weights. Yet, the research in the field of dynamic fire risk prediction model construction in stadiums is extremely lacking. In addition, the main working mode of the government and fire safety management departments is "man-sea warfare to carry out inspections and turns to filter key points" [7], and relies solely on human experience to subjectively judge whether the fire risk level is high, and the ability to actively detect and give an advance warning is weak. Therefore, it is necessary to establish a high-quality fire decision support system or model to make forward predictions of fires in stadiums, evaluate the possibility or risk level of fires, and discover and deal with high-risk hidden dangers in time.

With the continuous maturity of big data analysis technology [8] and ML intelligent algorithms, the application in building fire prediction seems very promising by analyzing massive historical data for forward-looking predictions. Most of the studies focus on predicting property damage [9], casualties [10], accident severity [11–13], and other ex post evaluation indicators, and they have achieved good results. Building a fire prediction system can be roughly divided into two aspects: community-level building fire prediction and property-level building fire prediction.

Community-level building fire prediction: Surya et al. [13] proposed a new framework for real estate fire risk prediction based on statistical machine learning. The research results show that the optimal model of the artificial neural network for evaluating the frequency of catastrophic fires can detect and predict the occurrence of fires in a timely and accurate manner. Liu et al. [12] proposed a cross-region transfer learning approach to identify fire hazard frameworks in communities such as parking lots, public spaces, and shopping malls. Dividing the community fire danger into nine grades, its recognition performance has been improved by 12%, 15%, 16%, 15%, and 15% in overall accuracy, precision, recall, F1 score, and AUC, respectively.

Property-level building fire prediction: This level refers to risk prediction studies that predict fire risk in terms of property damage and casualties. In Pittsburgh, Pennsylvania, Madaio [9] proposed a framework for building fire risk predictions. Models in this framework predict a construction fire score of 1–10 (lowest risk to highest risk) for building properties. In the experimental results, the recall value of the XGBoost model is 0.55, and the AUC value is 0.77. Anderson-Bell, J. [14] constructed a framework for predicting fire risk in buildings in London. It uses Fire Brigade incident data, aerial imagery, and a digital surface model (DSN). The final model presented achieved an ROC AUC of 0.8195 on the test set. Firebird [10] is a model for predicting building fire risk in Atlanta. It uses fire event data (time, location, and cause of fire), commercial property structure data, and property fire risk inspection data and predicts a fire risk score between 0 and 1 for building

properties. The evaluation result is that the random forest (RF) model performs the best with an AUC value of 0.8246. However, ML algorithms are not one-size-fits-all, and the results found in other studies cannot be used to determine that this classifier will perform best with minimal error on the Chinese stadium dynamic fire risk prediction dataset. Due to the variety of ML algorithms, the algorithm that is more suitable for dynamic fire risk prediction in stadiums has not been scientifically verified. It is necessary to carry out multi-dimensional experimental research on various classification algorithms and model testing methods.

In this paper, to further assist stadiums in fire supervision management and resource planning, a gradient boosting-recursive feature elimination (GB-RFE) method is designed to extract fire risk features and reduce the redundancy of features. Then, a multicollinearity test was performed on the optimized feature subset using Pearson correlation analysis to eliminate strongly correlated risk factors. The obtained optimal feature subset is used as the input data set of the ML algorithm for classification training. To avoid the influence and analysis of redundant features on the model prediction results, the model performance and operation efficiency can be improved. A fire risk prediction method based on k-fold cross-validation and gradient boosting decision tree fusion is proposed. Its main purpose is to allow the fire management department of the unit to gather fire resources to rectify or eliminate major fire hazards for the first time, and nip more fire hazards in the bud. The practical significance of the experimental results is that the use of the ML model can speed up some information analyses, and make them more objective and effective in predicting performance than based on human subjective analysis, providing a scientific basis for the prevention and management of stadium fire accidents. Summarily, the main contributions of this paper are:

- We propose a risk prediction model of a gradient boosting decision tree combined with the K-fold cross-validation strategy, which can effectively predict the fire risk level of stadiums based on dozens of factors. We show that with basic information about stadiums (fire acceptance status, fire host failure rate, stadium size, etc.), we can predict in advance the likelihood of a stadium fire in the future.
- We show that by using the GB-RFE method to screen and optimize the indicators, the optimized fire risk feature can replace all the features to represent the fire risk of the stadium, and its model performance also achieves the same or similar effect.
- With reference to standard regulations and related literature, we design threshold intervals from both static and dynamic aspects to quantify and classify fire risk assessment indicators.

The rest of this paper is organized as follows. Section 2 describes the dataset and stadium fire risk classification. Section 3 discusses the proposed method. In Section 4, the experimental results are analyzed in detail. Section 5 validates the optimal predictive model and discusses research limitations and scope for future work. Finally, conclusions are drawn in Section 6.

2. Dataset

The dataset we used was compiled by the Wuhan Fire Emergency Rescue Detachment on the fire risk source data of intelligent stadiums from the years of 2017 to 2020. In this study, 48 features were selected, where 47 were input attributes and 1 was an outcome or prediction attribute (i.e., stadium fire risk class). These attributes include building inherent safety, fire safety personnel management, fire facility equipment management, hazard management, unit fire data, building fire data, fire files, agencies, and personnel. The description and types of attributes are shown in Appendix A.

Stadium fire risk class is the predictive attribute that measures the risk level of casualties and property losses in the event of a fire in a stadium. According to Table 1, there are five stadium fire risk classes [6] as stratified by severity: Level I (not at risk), Level II (Low risk), Level III (Medium risk), Level IV (High risk), Level V (Extremely high risk). Figure 1 shows the distribution of risk levels for stadiums. To avoid over-fitting in the training

and testing phases, the five-fold cross-validation and Stratified five-fold cross-validation techniques were used to randomly divide the dataset into five equal-sized subsamples.

Table 1. Classification of fire risk ranks.

Risk Score	Risk Rank	Attribute Requirements
[90–100]	Level I (Not at risk)	Low priority
[80–90)	Level II (Low risk)	Regular inspection
[70–80)	Level III (Medium risk)	Frequent regular inspection and fire safety management.
[60–70)	Level IV (High risk)	The probability of fire accidents is extremely high, some casualties and particularly heavy property losses, take immediate measures.
<60	Level V (Extremely high risk)	The probability of a fire accident is extremely high, with a large number of casualties and particularly heavy property losses, take immediate measures.

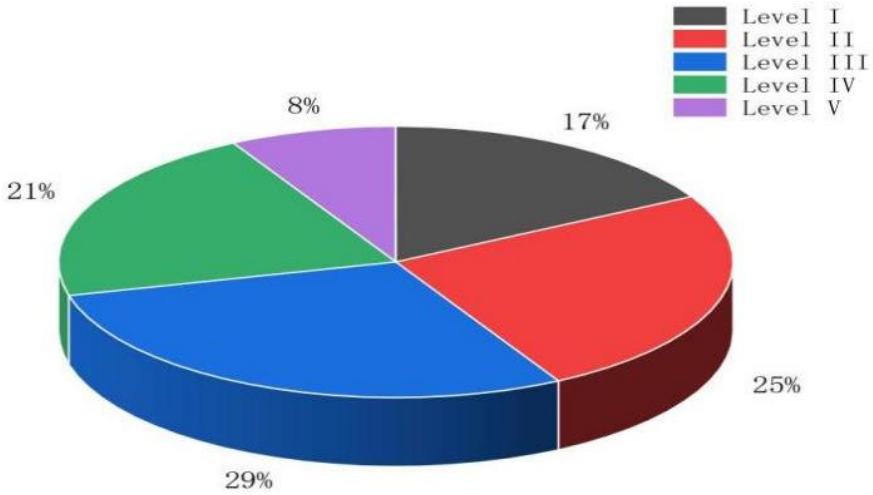


Figure 1. Distribution of fire risk levels for stadiums.

3. Methodology

To guarantee the quality of experimental results, in this study, we propose a data mining architecture that consists of three stages. Figure 2 shows an overview of the data mining architecture.

The goal of this study is to identify significant features and machine learning algorithms for building an optimal classification model to predict the level of fire risk in stadiums. During the data preparation phase, the quality of the dataset is assessed based on the percentage of missing values and is preprocessed to become a clean dataset (data cleaning). Next, since the unit measurement of various fire risk attributes is not uniform, the design index can quantify the threshold interval and perform the numerical quantification of the index (data conversion). In the modeling stage, through feature selection methods and correlation analysis, significant features were selected for experiments, and 12 risk prediction models were established by using two cross-validation techniques and six machine learning algorithms. The significant features were replaced with full features (47 features) and the same technical operations were performed to conduct model comparison studies. Finally, different indicators are used to measure the performance of the prediction model in the evaluation stage, and the risk prediction model with the best performance is decided. Sections 3.1–3.5 describes data preprocessing, feature selection, feature correlation analysis, classification modeling, and performance measurement in more detail.

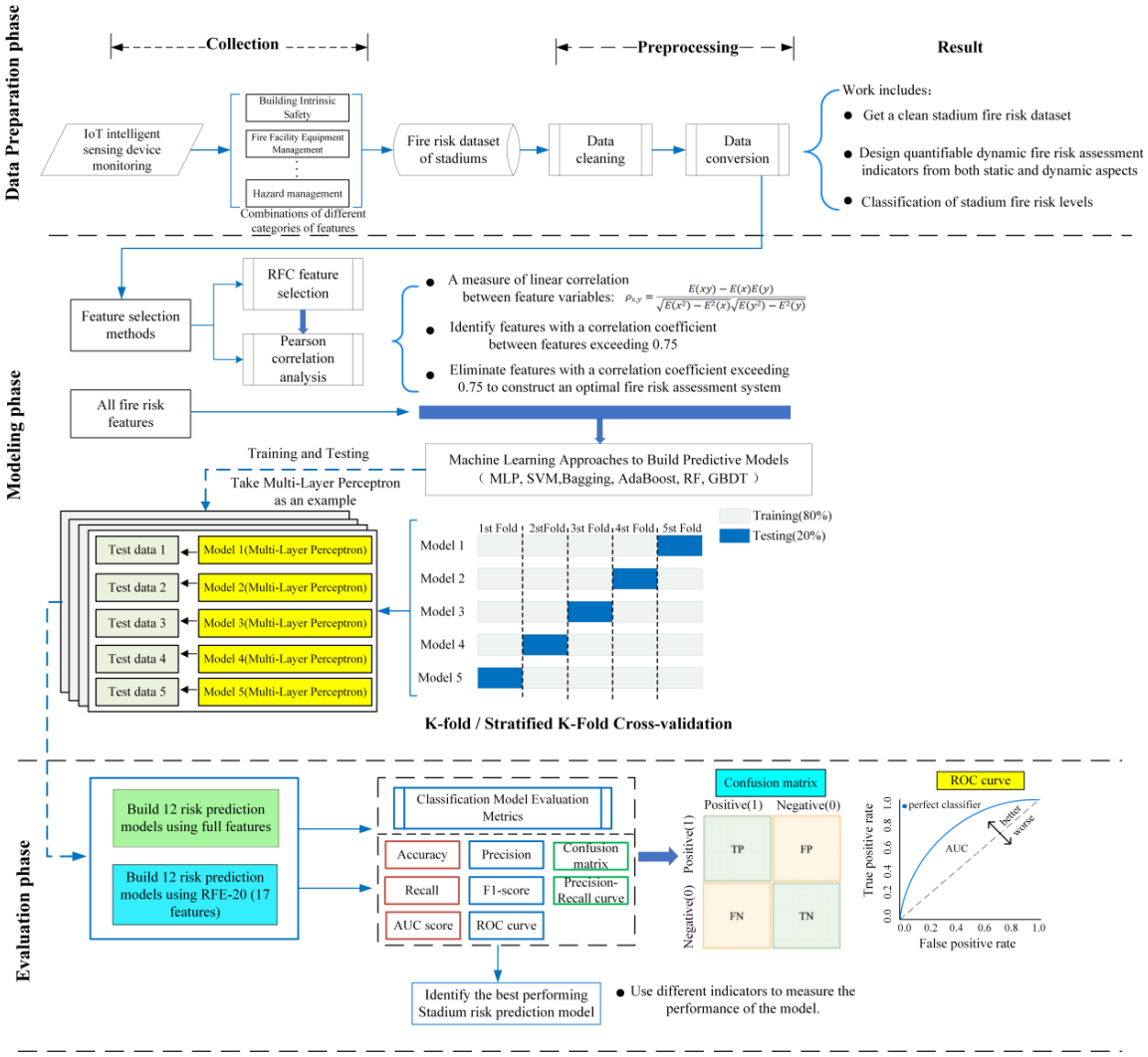


Figure 2. The main flow of the data mining method.

3.1. Data Preprocessing

In the context of IoT sensing devices transmitting data, the conditions for data collection are not perfect, where faulty IoT sensing devices (intermittent loss of sensor connectivity) or human oversight can result in noisy data containing errors, redundancy, and missing values set. Moreover, the unit measurement of various fire risk characteristics in the collected data is not uniform, which is not conducive to the construction of classification models. Clearly, the information extracted from "noisy data" (i.e., unreliable data) can be wrong, leading to a high probability that the day-to-day fire management decisions to be made are unsound [15]. So, all the above problems have to be dealt with in the preprocessing stage by applying various data augmentation methods like data cleaning and data transformation.

3.1.1. Data Cleaning

Stadium fire risk data is often missing due to various reasons such as IoT sensing equipment failure, human negligence, and technical problems in the IoT remote monitoring system and cloud servers. The missing degree of the collected data is lower than 30% of the set threshold [15], and the difference method is used to complete it. For discrete features, such as the liquid level of the fire pool water tank, the mean interpolation method is used to supplement the median of the features. Type features such as fire hydrant control cabinet status, sprinkler control cabinet status, and other indicators are quantified according to

design thresholds and supplemented with discrete feature types. Mean interpolation and mode interpolation are mainly used to predict and fill the rest of the data according to the feature to eliminate noise and correct inconsistency.

3.1.2. Data Conversion

Because the unit measurement of various fire hazard characteristics in the collected data set by the urban IoT monitoring system is not uniform, in order to realize the quantification and classification of prediction results, a quantifiable threshold interval is designed to classify and quantify various fire safety hazards.

The dynamic fire risk of stadiums usually needs to be comprehensively considered from both static and dynamic indicators. Static indicators are considered from both inside the building (fire resistance rating, evacuation facilities and firefighting facilities and equipment configuration, etc.) and outside the building (fire separation distance, fire lane, rescue site, etc.). Among them, whether the fire resistance level, fire protection facilities and equipment configuration, fire separation distance, etc., meet the design specifications or whether the performance-based design is reasonable can be characterized by the indicator of "pass fire protection acceptance". The sealing method of the stadium has a great influence on the fire and smoke exhaust, and the index of the sealing degree of the building structure of the stadium is selected. The number of seats in the venue reflects the building scale and potential inherent risks of the venue to a certain extent, and the number of seats in the venue is selected as an indicator. In summary, the static indicators include building structure, venue size, fire protection acceptance, etc. The threshold interval of venue size and fire protection acceptance are divided according to the standard specification "Uniform Standard for Civil Building Design (GB50352-2019)" and "Fire Protection Law of the People's Republic of China (2019)". The threshold interval of the building structure is divided by referring to the relevant literature [16]. The details are shown in Table 2.

Table 2. Examples of static fire risk indicators for stadiums.

First-Level Metrics	Second-Level Metrics	Third-Level Metrics	Level I [90–100]	Level II [80–90)	Level III [70–80)	Level IV [60–70)	Level V (<60)
Building Inherent Safety	Building fire performance	Venue size/seats	<3000	[3000,5000)	[5000,10,000)	[10,000,50,000)	≥50,000
		Building structure	Outdoor	Open	Partially Open	Enclosed	—
	Fire acceptance	Whether it has passed the fire inspection	Pass	—	—	Fail	—

In addition to the inherent safety attributes of buildings, other factors are used as dynamic disturbance factors, including three aspects: personnel management, facility equipment management and hidden danger management. Among them, personnel management mainly considers the on-the-job situation (system punching, video surveillance), training situation (number of licensees, uploading safety training records), and fire drills (uploading drill records and time calculation). The hidden danger management mainly considers the completion rate of inspection points, the stock of hidden dangers, the highest level of hidden dangers, and the rectification of hidden dangers. Facility and equipment management mainly considers the fire main engine, automatic sprinkler system, fire hydrant fire extinguishing system, fire door/fire shutter, smoke prevention and exhaust system, fire pool/water tank, and unit maintenance.

Taking the main engine of fire fighting equipment as an example, the threshold design is shown. Obtain the number of fault points and online points on the fire host through real-time monitoring, and design a quantifiable indicator "failure ratio". Failure ratio refers to the ratio of the number of fault points on the host and the number of online points. According to relevant literature, if the failure rate is 0%, it will be Level I (90–100 points), which is not at risk. If the failure ratio is (0%, 5%), it is a Level II (80–90 points), which belongs to low risk. The threshold interval of fire host status and fire host power detection

is divided according to "Maintenance and Management of Building Fire-fighting Facilities (GB25201-2010)" and "Code for Design of Automatic Fire Alarm System (GB50116-2013)". The threshold interval of the shielding ratio (number of masked points on the host/number of online points) is divided according to the relevant literature [1], and its threshold interval is shown in Table 3.

Table 3. Examples of dynamic fire risk indicators for stadiums.

First-Level Metrics	Second-Level Metrics	Third-Level Metrics	Level I [90–100]	Level II [80–90)	Level III [70–80)	Level IV [60–70)	Level V (<60)
Facility Equipment Management	Fire host	Fire host status	Normal	—	No data	Offline time ≤ 24 h	Offline time > 24 h
		Fire host power detection	Both the main and standby fire power supply signals are detected	—	One of the main and backup fire-fighting power supply signals is detected	—	The main and backup fire-fighting power signals are not detected
		Failure ratio	0%	(0%,5%]	(5%,10%]	(10%,20%]	>20%
		Shielding ratio	0%	(0%,5%]	(5%,10%]	(10%,20%]	>20%

3.2. Recursive Feature Elimination

After data preprocessing, due to the large number of dimensions of fire risk features in stadiums, it is necessary to select meaningful features and input them into machine learning algorithms for training. The feature selection method is a set of methods for selecting attribute variables closely related to the occurrence of a stadium fire. The feature selection method used is Recursive Feature Elimination (RFE) [17], which belongs to the Wrapper method and is a greedy algorithm for finding the optimal feature subset. It iteratively eliminates the one least-relevant feature (i.e., the lowest ranking criterion score), then repeats the process on the remaining features until all features are traversed. In this process, the order in which the features are eliminated is the sorting of the features, and finally the optimal feature subset is selected according to the sorting criterion score.

The stability of the RFE largely depends on the base estimator used for the iteration, and the Gradient Boosting algorithm is chosen as the base estimator, and the accuracy is used as the cross-validation score. The processed data is randomly divided into two parts: training set and test set. Using the basic idea of K-fold cross-validation and stratified K-fold cross-validation, the model parameter CV = 5 is selected. Features selected for this approach will be reported and discussed in Section 4.1.

3.3. Feature Correlation Analysis Based on Pearson

The Pearson correlation coefficient is a measure of the linear correlation between distributions. The output range is $[-1, 1]$, where 0 represents no correlation, positive values represent positive correlations, and negative values represent negative correlations. The closer the absolute value of the correlation coefficient is to 1, the stronger the correlation. The closer the correlation coefficient is to 0, the weaker the correlation. By calculating the Pearson correlation coefficient of the feature variable by formula (1) [18], it can be judged whether the selected feature is reasonable. If the absolute value of the correlation coefficient between variables is greater than 0.75, there may be a multicollinearity problem, indicating that the feature selection is unreasonable [19]. Otherwise, the feature selection is more reasonable. This study utilizes the obtained 139 sample instances to make a correlation matrix heatmap of salient features, which will be reported in detail in Section 4.2.

$$\rho_{x,y} = \frac{\operatorname{cov}(x,y)}{\sigma_x \sigma_y} = \frac{E\big((x-\mu_x)(y-\mu_y)\big)}{\sigma_x \sigma_y} = \frac{E(xy) - E(x)E(y)}{\sqrt{E(x^2) - E^2(x)}\sqrt{E(y^2) - E^2(y)}} \quad (1)$$

where $\operatorname{cov}(x,y)$ is the covariance between x and y, μ_x and μ_y are the mean of x and y, respectively, μ_x and μ_y are the standard deviation of the sum, and E is the mathematical expectation.

3.4. Performance Measure

3.4.1. Classification Metrics

In the field of classification, a key factor in evaluating the performance of any model is the ability to train the model to correctly classify the categories of stadium fire risks. Evaluation indicators are quantitative indicators of model performance. An evaluation metric can only reflect part of the performance of the model. If the selected evaluation indicators are unreasonable, wrong conclusions may be drawn. Therefore, different evaluation indicators should be selected for specific data and models. This study uses several commonly used metrics to evaluate the performance of the learned model, namely accuracy, precision, recall, macro F1 score, AUC score, ROC curve and precision-recall curve, most of which are calculated based on the confusion matrix. In Table 4, a confusion matrix is a representation of the predictions for a classification problem that summarizes the number of correct and incorrect predictions for each class.

Table 4. Classification confusion matrix.

	Predicted Positive	**Predicted Negative**
Actual Positive(P)	True Positive(TP)	False Negative(FN)
Actual Negative(N)	False Positive(FP)	True Negative(TN)

(1) Accuracy: The ratio of the correct number of samples classified by the classifier to the total number of samples.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{2}$$

(2) Precision: The ratio of the total number of positive samples correctly classified by the classifier to the total number of samples identified as positive samples by the classifier.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

(3) Recall or Sensitivity: The ratio of the total number of correct positive samples classified by the classifier to the total number of real positive samples.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4}$$

(4) F1-score: as the harmonic mean of recall and precision, it is better than independent precision or recall indicators, which is an important indicator for evaluating classification models [20]. Precision and recall have their own shortcomings. If the threshold is high, the precision is high, but there will be a lot of data loss; if the threshold is low, the recall will be high, but the prediction will be very inaccurate. Therefore, the F1-score is used to evaluate the classifier more comprehensively and can balance the effects of precision and recall.

$$\text{F}_1 = 2\frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{5}$$

In addition, the area under the ROC curve (AUC) and the area under precision-recall curves (Auprc) can be used as scalar metrics to evaluate classification performance.

3.4.2. Cross-Validation

Cross-validation is a statistical method for evaluating and validating data mining algorithms [21], which divides a dataset into two parts: one for training and the other for testing. In cross-validation, the training and testing sets must be crossed in consecutive rounds in order to run and validate the model in each cluster. There are many cross-validation methods, such as k-fold cross-validation (including k-fold, stratified k-fold), leave out one, and shuffle split. To ensure the consistency of the testing set and the original

distribution of the data, k-fold cross-validation and stratified k-fold are used, and the details will be reported and discussed in Sections 4.3.1 and 4.3.2.

1. K-fold cross-validation

To minimize the low performance associated with random dataset splits of training and testing data, we tend to use k-fold cross-validation. In k-fold, the entire dataset (S) is randomly divided into K equal-sized subsets (S_1, S_2, ... , S_k). The model is trained and tested k times, each time (t_1, t_2, ... , t_k) it is trained on all but one subset (S_t) and on the remaining single subset (S_t) is tested on. Finally, the average value of k evaluation indicators is used as the final result.

2. Stratified K-Fold cross-validation

Stratified K-fold cross-validation is a stratified sampling cross-validation method, which ensures that the proportion of each category of samples in the training set and testing set remains the same as the original data set. This ensures that a particular class does not appear in the validation or training dataset, especially if the dataset is imbalanced.

3.5. Classification Modeling Using Data Mining Algorithms

In this study, there are six widely used supervised machine learning algorithms for classification modeling, these algorithms include: Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), AdaBoost, Random Forest (RF), Gradient Boosting, and Bagging to get unbiased predictions. The experimental environment for model construction uses the Python 3.8 language and Pytorch 1.9.0 framework, and is carried out on an NVIDIA QUadro T2000 graphics card. In the training phase, the training batch size is 20, the training epochs is 50, and the initial learning rate is 0.01. The optimizers we use are Stochastic Gradient Descent (SGD) and grid search methods to optimize hyperparameters. A classification model for predicting the dynamic fire risk degree of stadiums is established through experiments (see Figure 3):

- Using full features (47 features), combined with two cross-checks and six machine learning algorithms to build 12 risk prediction models.
- Selecting a significant feature subset (17 features) of recursive feature elimination, and using two cross-validation methods and six machine learning algorithms to establish 12 risk prediction models.

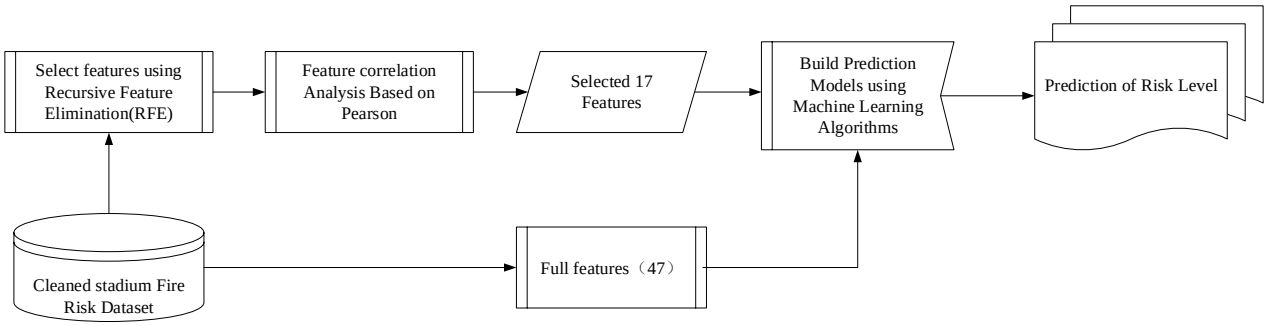


Figure 3. Risk prediction model developed using significant features and machine learning algorithms.

4. Experimental Results

4.1. Selected Features

In the process of feature selection, irrelevant features are removed, and significant features are selected to obtain unbiased optimal results. RFE automatically adjusts the number of selected features through cross-validation, and selects the best-performing subset of features for risk prediction. It can be seen from Figure 4 that the optimal number of features is 17, and these features are: FRBMaterials, Evacuation signs, Rescue field, NO_CertificatesPersonnel, Fire drills, Fire host, rFHF, nrWaterPressure, PPCSSmoke, CC-

Smoke, lFP, UBMCompany, ULMTime, rIPC, rHD, BgLayout, and FSupervisor. These features are the following types of firefighting fire information:

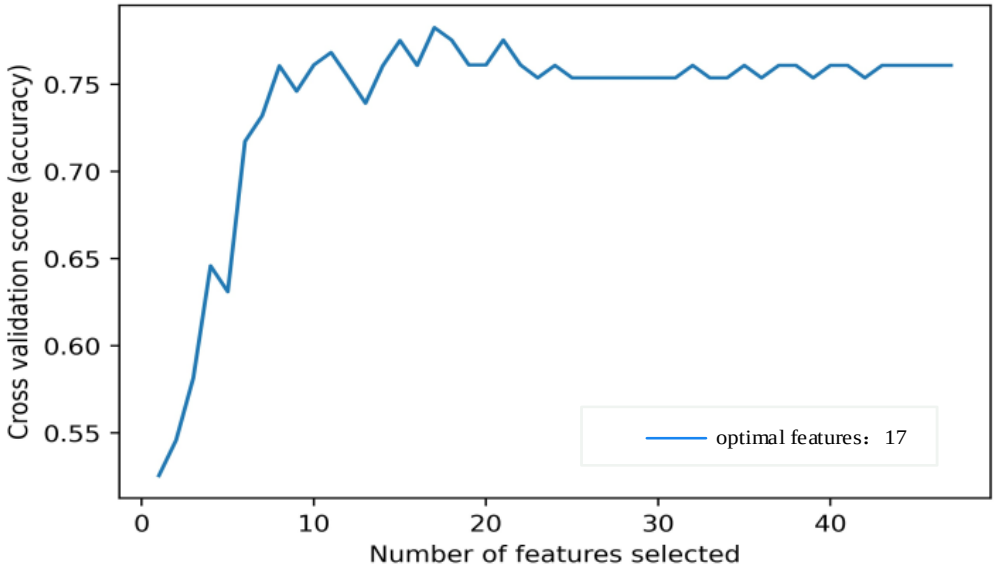


Figure 4. Cross-validation scores under different numbers.

1. Fire acceptance: FRBMaterials, Evacuation signs, Rescue field;
2. Fire Safety Personnel Management: NO_CertificatesPersonnel, Fire drills;
3. Fire Facility Equipment Management: Fire host, rFHF, nrWaterPressure, PPCSSmoke, CCSmoke, lFP, UBMCompany, ULMTime;
4. Hazard management: rIPC, rHD;
5. Unit fire data maintenance: BgLayout, FSupervisor.

4.2. Feature Correlation Analysis

As shown in Figure 5, the correlation coefficient of the same feature on the diagonal is 1, and the closer the color of the matrix diagram is to black, the stronger the feature correlation and the positive correlation. The closer the color of the matrix diagram is to green, the stronger the feature correlation and negative correlation. The absolute value of the correlation coefficient between most features is less than 0.75. Exceptionally, the correlation coefficients of "FRBMaterials", "Evacuation signs", "rIPC", and "Fire drills" are all greater than 0.75. Since rIPC, IFP, FRBMaterials, Rescue sites, and Evacuation signs are all significant indicators that affect the fire risk of stadiums, after detailed consideration and consultation with fire experts, these properties are preserved. The merged dataset contains 139 observations and 17 attributes as the final cleaned dataset.

4.3. Cross-Validation

This section presents the results obtained by cross-validation techniques using RFE-Top20 (17 features): K-fold and stratified K-fold.

4.3.1. K-Fold

We split our dataset into five clusters using K-fold, each cluster giving a different accuracy. Table 5 and Figure 6 show the accuracy for each cluster and the mean using K-fold cross-validation. Figure 7 shows other performance metrics of the six classifiers using K-fold cross-validation.

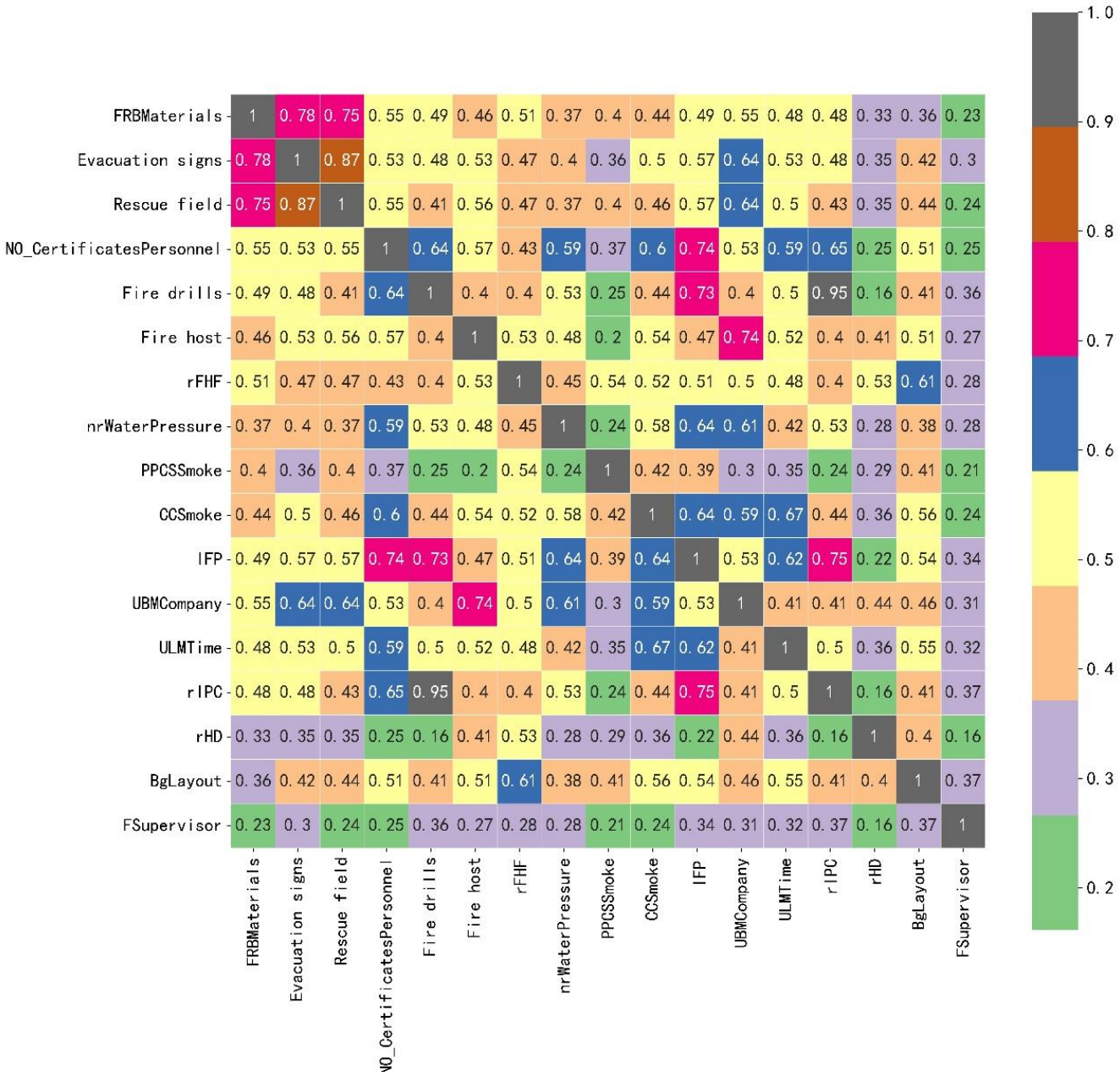


Figure 5. Correlation matrix heatmap for selected features.

Table 5. K-fold Cross-Validation Accuracy Prediction For Six Classifiers.

	Cluster 1 (%)	Cluster 2 (%)	Cluster 3 (%)	Cluster 4 (%)	Cluster 5 (%)	Average (%)
MLP	81.42	81.42	82.85	87.85	84.44	83.59
SVM	85.71	85.71	91.42	82.85	88.14	86.76
RF	84.28	94.28	95.71	97.14	89.62	92.20
Bagging	91.42	94.28	95.71	92.85	86.66	92.18
AdaBoost	90.00	94.28	94.28	91.42	89.62	91.92
Gradient Boosting	92.85	94.28	97.14	95.71	86.66	93.32

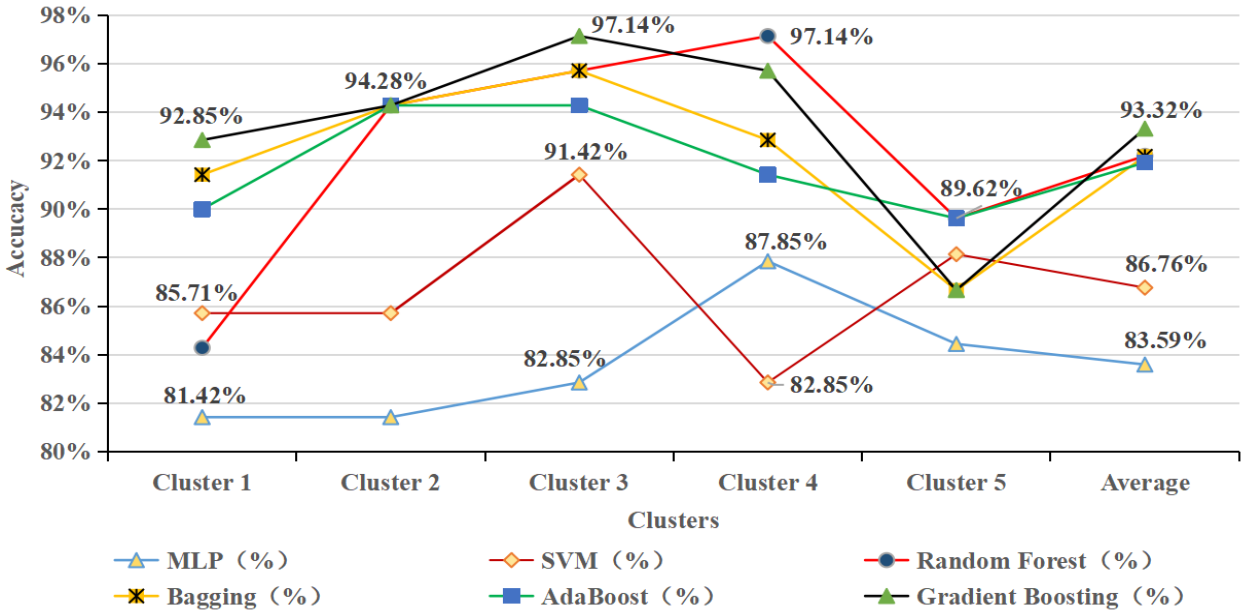


Figure 6. K-fold Cross-Validation For Six Classifiers.

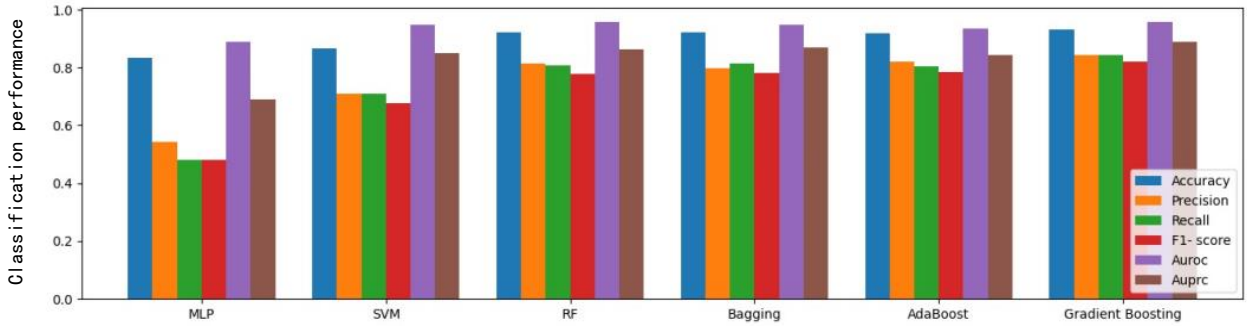


Figure 7. Other performance metrics for six classifiers using K-fold cross-validation.

4.3.2. Stratified K-Fold

We divided our dataset into five clusters using stratified K-fold, making sure that the proportion of each category in the training set and test set remained the same as the original data set, each cluster giving a different accuracy. Table 6 and Figure 8 show the accuracy for each cluster and the mean using stratified K-fold cross-validation. Figure 9 shows other performance metrics of the six classifiers using stratified K-fold cross-validation.

Table 6. Stratified K-fold Cross-Validation Accuracy Prediction For Six Classifiers.

	Cluster 1 (%)	Cluster 2 (%)	Cluster 3 (%)	Cluster 4 (%)	Cluster 5 (%)	Average (%)
MLP	79.28	86.42	88.57	90.00	93.33	87.52
SVM	90.00	88.57	91.42	92.85	88.14	90.19
RF	91.42	91.42	88.57	95.71	92.59	91.94
Bagging	90.00	87.14	88.57	94.28	92.59	90.51
AdaBoost	90.00	84.28	90.00	91.42	92.59	89.65
Gradient Boosting	91.42	87.14	88.57	94.28	92.59	90.80

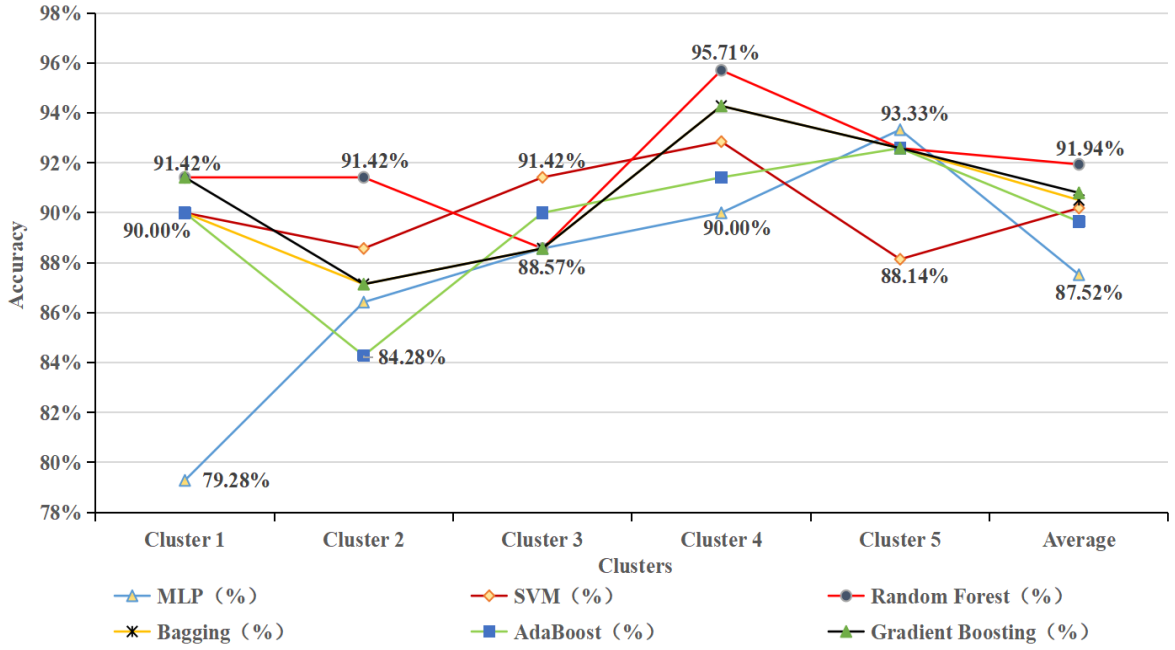


Figure 8. Stratified K-Fold Cross-Validation For Six Classifiers.

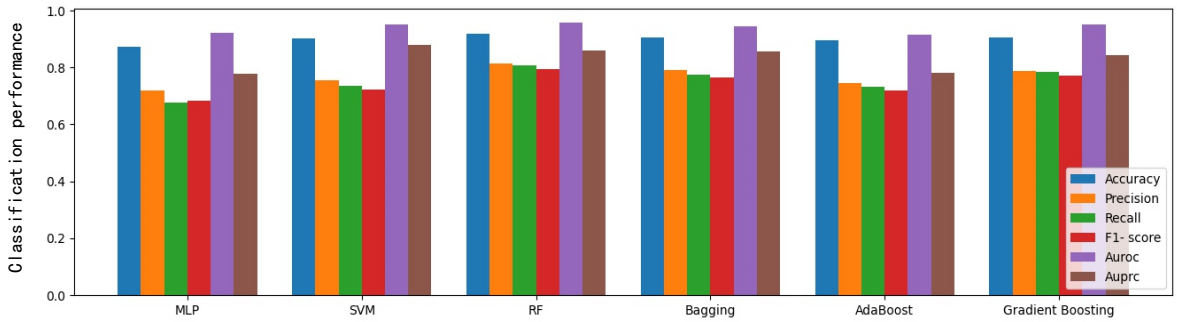


Figure 9. Other performance metrics for six classifiers using stratified K-fold cross-validation.

4.4. Performance of Classification Models

Using the stadium fire risk dataset, through the five-fold cross-validation technique and the stratified five-fold cross-validation technique, the performance of a classification model for predicting the fire risk level of a stadium is established and evaluated. It can be seen from Table 7a,b that all models achieve an accuracy of more than 71% and an accuracy of more than 83%. The Gradient Boosting model built with RFE-Top20 (17 features) achieves the highest accuracy (93.2%) and precision (84.2%) through the K-fold cross-validation technique. Moreover, through the K-fold cross-validation technique, the model built with full features also achieved more than 84.2% accuracy and precision. Furthermore, through the stratified K-fold cross-validation technique, the models built using RFE-Top20 and full features both achieved an accuracy higher than 86.0% in risk prediction with a precision ranging from 68.8% to 81.5%.

Table 7c,d shows recall and F1-score. Among all models, the Gradient Boosting model with full features and RFE-Top20 achieved the highest recall (84.3%) and F1-score (81.9%) by the k-fold cross-validation technique. On the other hand, in Table 7e, the auroc ranged from 90.1% to 96.2%, where the Gradient Boosting model with selected full features was the optimal model to distinguish extremely high risk, high risk, medium risk, low risk, and not at risk categories. As shown in Table 7f, most of the models obtained a ratio of over 84.0% in the auprc metric, among which the highest auprc (89.8%) was obtained by the Gradient Boosting model with full features under the K-fold cross-validation technique. Overall, the

models developed using full features performed well in distinguishing extremely high risk, high risk, medium risk, low risk, and not at risk classes, with auroc ranging from 90.1% to 96.2% while those developed using RFE-Top20 features models show similar auroc (i.e., range from 88.8% to 95.9%).

Table 7. Accuracy, precision, recall, F1-score, auroc, and auprc using the full features and selected features on the stadium fire dataset using five-fold cross-validation.

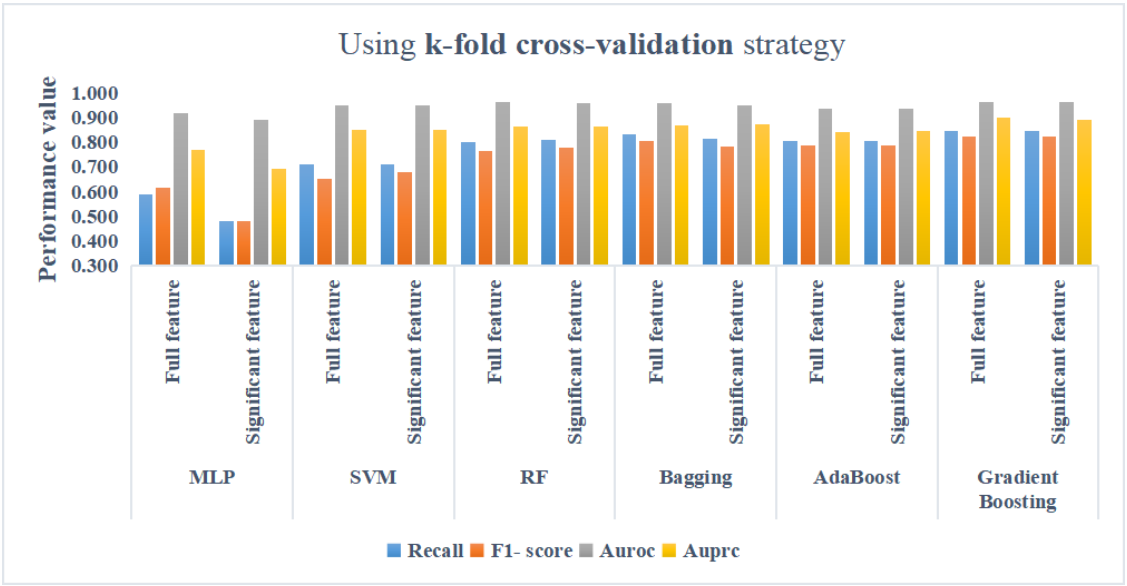
Performance Metrics	Machine Learning Algorithms	K-Fold Cross-Validation		Stratified K-Fold Cross-Validation	
		Full Features (47 Features)	RFE-Top20 (17 Features)	Full Features (47 Features)	RFE-Top20 (17 Features)
(a) Accuracy	MLP	87.1	83.5	86.0	87.4
	SVM	86.7	86.7	90.1	90.1
	RF	91.6	92.1	91.3	91.9
	Bagging	92.4	92.1	91.8	90.4
	AdaBoost	91.8	91.8	90.2	89.6
	Gradient Boosting	93.2	93.2	91.6	90.7
(b) Precision	MLP	68.5	54.4	68.8	71.9
	SVM	71.0	71.0	75.4	75.4
	RF	77.5	81.5	80.2	81.5
	Bagging	82.5	79.8	81.0	79.0
	AdaBoost	81.9	81.9	77.6	74.7
	Gradient Boosting	84.2	84.2	80.7	78.8
(c) Recall	MLP	58.8	48.0	63.1	67.7
	SVM	70.8	70.8	73.4	73.4
	RF	80.0	80.8	79.5	80.9
	Bagging	82.9	81.4	80.0	77.5
	AdaBoost	80.5	80.5	77.3	73.3
	Gradient Boosting	84.3	84.3	80.4	78.6
(d) F1- score	MLP	61.3	48.0	64.1	68.4
	SVM	65.2	67.7	72.1	72.1
	RF	76.4	77.8	77.8	79.6
	Bagging	80.1	78.0	79.0	76.5
	AdaBoost	78.6	78.4	77.3	71.9
	Gradient Boosting	81.9	81.9	78.7	77.0
(e) Auroc	MLP	91.4	88.8	90.1	92.1
	SVM	94.8	94.8	95.4	95.3
	RF	95.9	95.8	96.1	95.9
	Bagging	95.7	94.9	94.1	94.5
	AdaBoost	93.4	93.5	91.9	91.7
	Gradient Boosting	96.2	95.9	95.1	95.1
(f) Auprc	MLP	76.6	68.9	75.1	77.7
	SVM	84.8	84.9	88.2	87.9
	RF	86.2	86.2	87.7	86.0
	Bagging	86.8	87.1	84.6	85.7
	AdaBoost	84.0	84.2	78.2	78.2
	Gradient Boosting	89.8	88.8	84.4	84.4

5. Discussion and Future Work

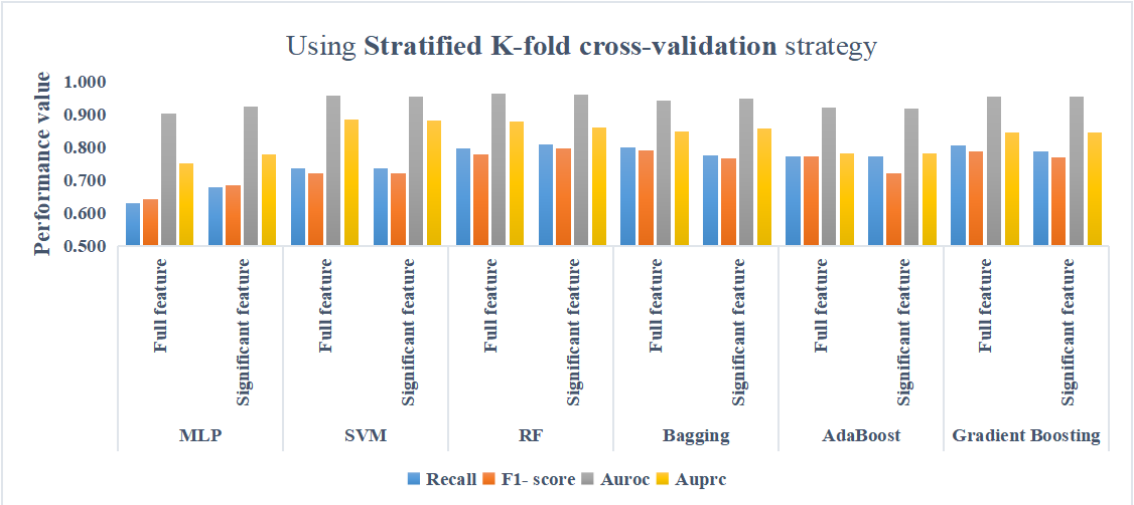
5.1. Comparison of Performance Metrics between Predictive models Using Significant Features and Full Features

To verify that full features are replaced by the significant features selected by the feature selection method, and for the dynamic fire risk of the stadium to be effectively characterized, in this experiment, a set of six prediction algorithms, MLP, SVM, RF, Bagging, Adaboost and GBDT, for the prediction of fire risk level were used. Using two cross-validation strategies (choose CV = 5), we can calculate the average performance metrics for

each of the classifiers, namely recall, F1-score, auroc and auprc. Figure 10a,b shows that, in the case of stratified k-fold cross-validation, based on recall, F1-score, auroc and auprc, all models using significant features are almost better than full features. In the case of k-fold cross-validation, the performance of all models using significant features is slightly lower than that of full features, such as in the MLP model, the recall of significant features is 48.0%, while the recall of full features is 58.8%. The F1 score of significant features is 48.0%, while the F1 score of full features is 61.3%. In addition, in the Gradient Boosting model, the auprc of the full feature is 89.8%, which is slightly higher than the auprc of the significant feature of 88.8%. These results indicate that models with significant features may perform better than or similarly to models with full features. Moreover, most models outperformed full features on selected features. This suggests that the implementation of feature selection methods is worthwhile for improving the performance of risk prediction models. Removing redundant features can reduce model processing time and complexity, while also improving model quality.



(**a**)



(**b**)

Figure 10. Performance analysis of the classifiers using (**a**) Using K-fold cross-validation strategy; (**b**) Using stratified K-fold cross-validation strategy.

5.2. Optimal Risk Prediction Model

5.2.1. Performance of Risk Prediction Models

In building predictive models based on full features and RFE-Top20 features, overall, the accuracy of all models in RFE-Top20 using the K-fold cross-validation technique (83.5% to 93.2%) was higher than that of RFE-Top20 with the Stratified K-fold cross-validation technique (87.4% to 90.7%). The precision (ranging from 54.4% to 84.2%) of all models of RFE-Top20 using the stratified K-fold cross-validation technology is almost similar to that of RFE-Top20 using the K-fold cross-validation technology (ranging from 68.5% to 84.2%).

However, accuracy and precision do not guarantee that the performance results obtained are acceptable, and it may be biased towards the dominant class due to dataset imbalance. Due to the uneven distribution of classes, stratified K-fold cross-validation is used to overcome this problem [22], and five different performance evaluation indexes (accuracy, precision, recall rate, F1-score, AUC score) are used to compare the performance evaluation of classification models. The name of each model in Table 8 is represented by the cross-validation type, ML, and feature set combination. We count the frequency of the top five models for each performance metric. Table 9 lists the top three models in the six performance metrics. According to Table 9, using the stadium fire risk dataset for evaluation, Gradient Boosting + REF-Top20 + K-fold and Gradient Boosting + Full features + K-fold are both identified as one of the top five models for all six performance metrics. Meanwhile, the frequency ratio of the two models is 6:6. In terms of frequencies shown in Table 9, two models developed using Gradient Boosting have become the top models for this dataset, making Gradient Boosting the best performing ML algorithm in this study.

The F1-score can be used as a metric to evaluate a classification model, which reflects the overall performance of the classification model. Through K-fold cross-validation, the F1-score of models developed using REF-Top20 features ranged from 61.3% to 81.9%. The F1-score of models developed with full features ranges from 48.0% to 81.9%. Based on F1-score, through the K-fold cross-validation technique, the model developed using Gradient Boosting of REF-Top20 features achieved the highest F1-score (81.9%), which determined it to be the best performing model.

Table 8. Performance Analysis of Six Risk Prediction Classification Models.

Performance Metrics	Stadium Fire Risk Dataset	
	Model	**Value (%)**
(a) Accuracy	Gradient Boosting + REF-Top20 + K-fold	93.2
	Gradient Boosting + Full features + K-fold	93.2
	Bagging + Full features + K-fold	92.4
	RF + Full features + K-fold	92.1
	Bagging + REF-Top20 + K-fold	92.1
(b) Precision	Gradient Boosting + REF-Top20 + K-fold	84.2
	Gradient Boosting + Full features + K-fold	84.2
	Bagging + Full features + K-fold	82.5
	Adaboost + REF-Top20 + K-fold	81.9
	Adaboost + Full features + K-fold	81.9
(c) Recall	Gradient Boosting + REF-Top20 + K-fold	84.3
	Gradient Boosting + Full features + K-fold	84.3
	Bagging + Full features + K-fold	82.9
	Bagging + REF-Top20 + K-fold	81.4
	RF + REF-Top20 + Stratified K-fold	80.9
(d) F1-score	Gradient Boosting + REF-Top20 + K-fold	81.9
	Gradient Boosting + Full features + K-fold	81.9
	Bagging + Full features + K-fold	80.1
	RF + REF-Top20 + Stratified K-fold	79.6
	Bagging + Full features + Stratified K-fold	79.0

Table 8. *Cont.*

Performance Metrics	Stadium Fire Risk Dataset	
	Model	**Value (%)**
(e) Auroc	Gradient Boosting + Full features + K-fold	96.2
	RF + Full features + Stratified K-fold	96.1
	RF + Full features + K-fold	95.9
	Gradient Boosting + REF-Top20 + K-fold	95.9
	RF + REF-Top20 + Stratified K-fold	95.9
(f) Auprc	Gradient Boosting + Full features + K-fold	89.8
	Gradient Boosting + REF-Top20 + K-fold	88.8
	SVM + Full features + Stratified K-fold	88.2
	SVM + REF-Top20 + Stratified K-fold	87.9
	RF + Full features + Stratified K-fold	87.7

Table 9. The top three models appearing in the top five in all performance evaluation metrics.

Dataset	ML + Feature Combination + Cross-Validation	Frequency
Stadium Fire Risk Data	Gradient Boosting + REF-Top20 + K-fold	6
	Gradient Boosting + Full features + K-fold	6
	RF + REF-Top20 + Stratified K-fold	
	Bagging + Full features + Stratified K-fold	
	Bagging + Full features + K-fold	4

We present the confusion matrix for the predictions of the best performing model on the test set in Figure 11. For a confusion matrix, all values from the upper left corner to the lower right corner of the diagonal are correctly classified data samples, where the sum of each row (from left to right) is the number of that class in the total sample. For example, in the third row, we have 41 samples belonging to "Classes III", of which our model correctly predicted 40 (97.5%). Similarly, "Classes IV" and "Classes V" have 96.5% and 90.9% accuracy of correct predictions, respectively. We observe "Classes I", the model performance is not ideal. With 14 samples correctly classified and nearly one-third of the data samples misclassified into different classes, it showed an accuracy of 60.8%. Obviously, the model is overfitting on "Classes I", which means that the sample data for "Classes I" in training was not enough. Overall, the classification effect of the gradient boosting model is good, and it can solve the problem of misdiagnosis to a certain extent.

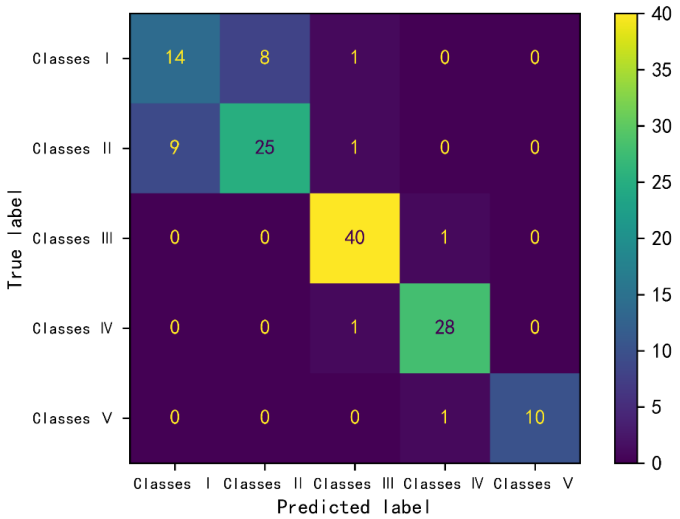


Figure 11. Confusion matrix using gradient boosting of significant features under K-fold cross-validation strategy.

The AUC score can also be used as a metric to evaluate the performance of a classification model, as this metric is useful and informative for evaluating a model's ability to recognize different classes. Figure 12 shows the ROC curves of six machine learning algorithm classification models, most of which have AUC scores over 0.90, SVM and GBDT have the highest AUC score (0.94), and the corresponding AUC score of MLP classification model (0.88) below 0.90. Overall, the six prediction models are relatively stable and effective.

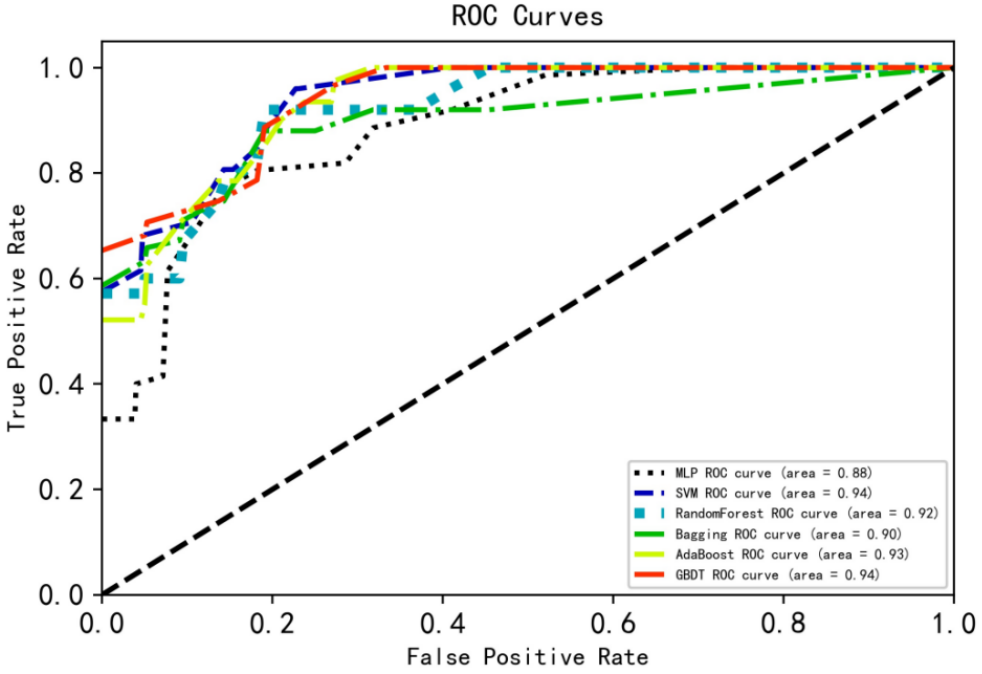


Figure 12. ROC curves for six machine learning algorithms using REF-Top20 by k-fold cross-validation.

However, the ROC curve more reflects the model's ability to correctly predict and rank positive and negative samples, but does not consider whether the distribution of positive and negative samples in the test data is balanced. Even though the distribution of the test data samples may change over time, the AUC value of the model will not be greatly changed (since the ROC curve is not sensitive), and tends to stabilize at a value. According to Figure 1, there are many instances of Class II and Class III samples in the stadium fire risk dataset, while there are few instances of Class V samples. Even if there is a category imbalance in the test samples, the model prediction results are good on the surface. Since the precision-recall curve is quite sensitive to the sample ratio, it can reflect the actual performance of the classifier as the sample ratio changes. Therefore, we also incorporate the precision-recall curve to further evaluate the prediction effect of the classification algorithm on the scenarios we apply. Figure 13 shows the precision-recall curves of the six machine learning algorithm classification models, AdaBoost obtains the highest AUC score (0.78), followed by SVM (0.77), and the model with the lowest AUC score is Gradient boosted decision tree (GBDT) (0.55). It can be seen from the results that the actual prediction effect of the GBDT model under the ROC curve is not ideal in the precision-recall curve, and the Adaboost and SVM models are more practical in this unbalanced dataset. With more balanced data, using AdaBoost or SVM may greatly improve model performance.

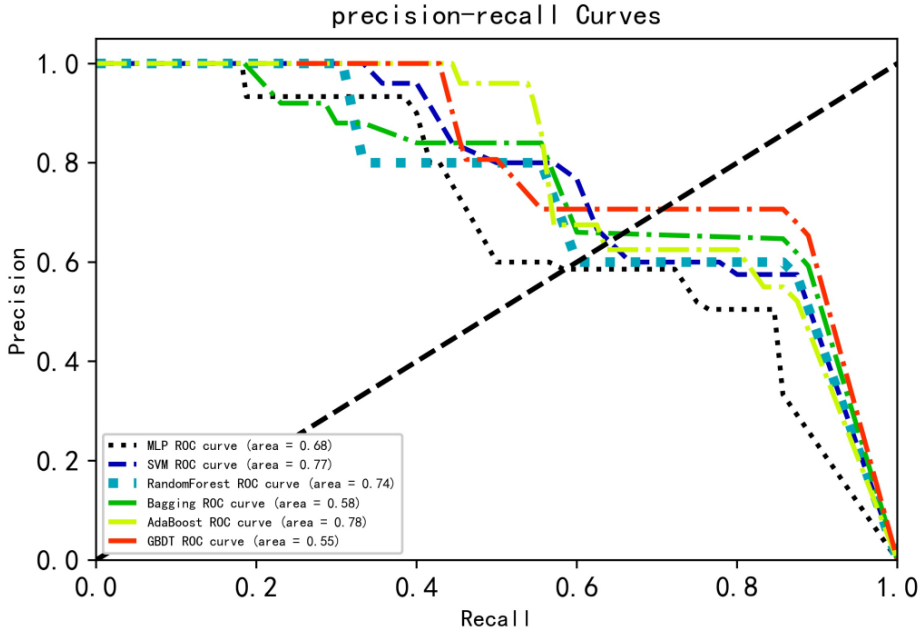


Figure 13. Precision-recall curves for six machine learning algorithms using REF-Top20 by k-fold cross-validation.

5.2.2. Comparison with Other Study

This section shows a performance comparison of the proposed model with existing research in predicting building fire risk. Table 10 shows the performance test of the model proposed in this study with the results obtained by existing research. Most of the existing studies only report accuracy results, and less for F1-score and AUC (including auroc and auprc). According to Table 10, the proposed model outperforms existing related studies. In addition to gradient boosting incorporating k-fold cross-validation, the top three top models reported in Section 5.2.1 have achieved over 92% accuracy and over 81% F1-score, and over 90% AUC score. This performance test demonstrates that the prediction model proposed in this study is acceptable in terms of risk prediction compared to existing research on building fire risk prediction.

Table 10. Comparison of the performance achieved by the proposed model with existing research.

Souce	ML Algorithm Used	Accuracy	Recall	F1-Score	Auroc	Auprc
Kim et al. [11]	Deep Neural Network	75.1%	—	—	—	—
Liu et al. [12]	TrAdaBoost (a typical transfer learning method)	89.0%	**88.0%**	—	89.0%	—
Poh et al. [23]	SVM	78.0%	—	—	—	—
Guan et al. [24]	K-nearest Neighbor	92.4%	—	—	—	—
Gholizadeh et al. [25]	AdaBoost (CART)	71.0%	—	69.0%	—	—
Dang et al. [26]	XGBoost (Test with balanced data)	91.0%	—	—	—	—
Zhu et al. [27]	Logistic Regression	—	80.3%	78.3%	—	—
Pirklbauer et al. [28]	Random Forest	—	—	—	91.0%	—
Wang et al. [29]	Neural Networks	—	55.8%	40.0%	76.3%	—
Zhang et al. [30]	Random Forest	91.2%	—	—	—	—
Chang et al. [31]	Neural Networks	89.1%	59.3%	70.1%	—	—
Proposed model	Gradient boosting with RFE-Top20 features using K-fold Cross-Validation	**93.2%**	84.3%	**81.9%**	**95.9%**	**88.8%**

5.3. Limitations and Future Work

This study also encountered some limitations. First, the dataset categories are not balanced. The presence of minority class labels (Class V) in the stadium fire risk dataset has not been practically addressed. To handle training data with minority class labels, it is common to resort to the widely adopted solution of resampling minority class instances. However, no matter how effective the resampling process is, it can seriously alter the original distribution of the data (if our goal is to balance the training set). The main reason is that during the sampling process, a large number of potentially very useful majority class label (Class II, Class III) instances may be inevitably removed from the training set, which reduces the generalization of the model. In this work, we attempt to partially ameliorate the above-mentioned data class imbalance importance problem (stratified sampling cross-validation scheme in Section 3.4.2) by evaluating the trained model on a validation set that is closer (in terms of size) to the initial data size, aiming to preserve as much of the initial data distribution as possible in this way. While some positive signs regarding the value of such a process have been identified (see Section 4.4), overall it did not significantly improve the model's effectiveness, and potential improvements need to be further investigated and planned in our future work.

Second, the dataset is relatively single. The fusion of large amounts of other types of data (e.g., fire images, time series observations) may improve model performance. Third, due to the scarcity of data, the stadium assessment only refers to the fire risk, which obviously affects the usefulness of the model. If more accurate property loss, casualty, social impact and other indicator data can be incorporated into the prediction model, the application value of this method will be greater. Furthermore, in future work, as experts acquire data, methods are needed to reduce the subjectivity of expert opinions. Moreover, deep neural networks and deep learning methods can be used and compared with machine learning methods. For instance, Zhang et al. [30] proposed a Deep Belief Network (DBN) with Recurrent LSTM Neural Network (R-LSTM-NN) for predicting fire hazard values in smart cities, which has an accuracy of 98.4% that is higher than our optimal model of 93.2%. With more data, the performance of deep learning methods may improve significantly.

6. Conclusions

In this study, according to the characteristics of IoT monitoring data, a quantifiable threshold interval is designed from both static and dynamic aspects, and the indicators of different data types are quantified and classified to obtain a quantitative data set. A classification model with significant features and full features and a machine learning algorithm based on stadium fire risk classification are developed to predict the risk level. In the experiment, the Wuhan Emergency Rescue Detachment was used to provide a real stadium fire risk data set, and a gradient boosting-recursive feature elimination (GB-RFE) method was designed to extract significant features. Briefly, there are 17 features, including FRBMaterials, Evacuation signs, Rescue field, NO_CertificatesPersonnel, Fire drills, Fire host, rFHF, nrWaterPressure, PPCSSmoke, CCSmoke, lFP, UBMCompany, ULMTime, rIPC, rHD, BgLayout, and FSupervisor, which are considered to be the most significant attributes for predicting dynamic fire risk levels in stadiums. The experimental results show that, based on the F1-score, the model developed using 17 significant features combined with K-fold cross-validation and Gradient Boosting obtained the highest F1-score (81.9%) and was identified as the best performing prediction model. In terms of AUC scores (From the perspective of ROC curve and precision-recall curve), AdaBoost achieves the highest auprc score (0.78), followed by SVM (0.77), and the model with the lowest auprc score is GBDT (0.55). Adaboost and SVM models have more stable performance and practical significance under this imbalanced dataset. Finally, the cloud platform based on China's Internet of Things big data to obtain a large amount of data from stadiums in a broader sense is applied to the future research of this machine learning method.

Author Contributions: Z.Z. and X.J. designed this research and collected the data set for the experiment. Furthermore, Y.L. developed the proposed methodology. X.F. wrote this manuscript and made the original draft. Y.L. and X.F. analyzed the data to show the validity of this paper and performed all the research steps. All authors have read and agreed to the published version of the manuscript.

Funding: This project was supported by the special project of safety production of Hubei emergency management department (No.KJZX201907011), the Youth project of Hubei Natural Science Foundation (No.2018CFB186) and the National Natural Science Foundation of China (No.51874213).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This study thanks Z.Z. for collecting the experimental dataset. In addition, X.F. wrote and produced the manuscript as well as analyzed the data to demonstrate the validity of this paper and performed all research steps.

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

Appendix A

Table A1. Description of attributes from Stadium Fire Risk Dataset.

Attribute Name	Description	Data Type and Value
Building Intrinsic Safety:		
Fire acceptance	Whether to pass the fire acceptance	Nominal–Pass/Fail
FRBMaterials	Flammability rating of building materials	Nominal–Pass/Fail
FEL	Fire emergency lighting	Nominal–Pass/Fail
Evacuation signs	Evacuation signs	Nominal–Pass/Fail
Fire lanes	Fire lanes	Nominal–Pass/Fail
Rescue field	Rescue field	Nominal–Pass/Fail
Rescue entrance	Rescue entrance	Nominal–Pass/Fail
BgStructure	Building structure	Nominal-Outdoor, open, partially open, enclosed
VeSize	Venue size	Nominal-Seats < 3000, 3000 $\leq$ Seats < 5000, 5000 $\leq$ Seats < 10,000, 10,000 $\leq$ Seats < 50,000, Seats $\geq$ 50,000
Fire Safety Personnel Management:		
NO_FSPR	Number of fire station personnel recorded	Nominal-Numbers $\geq$ 6, Numbers = 5, Numbers = 4, Numbers = 3, Numbers $\leq$ 2
FCRStaff	Staff in the fire control room	Numerical-%
FSTra	Fire safety training	Nominal-days $\leq$ 180, 180 < days $\leq$ 365, days > 365
NO_CertificatesPersonnel	Number of certificates of fire control room personnel	Nominal-Numbers $\geq$ 2, Numbers = 1, Numbers = 0
Fire drills	Fire drills	Nominal-days $\leq$ 180, 180 < days $\leq$ 365, days > 365
Fire Facility Equipment Management:		
Fire host	Fire host status	Nominal-Normal, no data, offline duration ($\leq$24 h), offline duration (>24 h)
FSPDetection	Fire host power detection	Nominal-Both are normal/ a normal/Neither is normal
rFHF	Fire host Failure ratio	Numerical-%
rFHS	Fire host shielding ratio	Numerical-%
rFAI	Fire alarm integrity ratio	Numerical-%
CCSprinkler	Sprinkler control cabinet status	Nominal-Automatic/ manual/offline/disconnected

Table A1. *Cont.*

Attribute Name	Description	Data Type and Value
nrWaterPressure	The normal rate of water pressure at the end of the sprinkler system	Numerical-%
WPFH	Worst point fire hydrant water pressure	Numerical-MPa
CCFireHydrantPump	Fire hydrant pump control cabinet status	Nominal-Automatic/ manual/offline/disconnected
rFDOI	Fire door operating integrity ratio	Numerical-%
rFSR	Fire shutter running integrity ratio	Numerical-%
PPCSSmoke	Smoke prevention power connection status	Nominal-connected/disconnected
CCSmoke	Smoke control cabinet status	Nominal-Automatic/ manual/offline/disconnected
lFWT	Fire water tank level	Numerical-mm
lFP	Fire pool level	Numerical-mm
UBMCompany	Unit-bound maintenance company	Nominal-Yes/No
ULMTime	Unit's Latest Maintenance Time	Nominal-days $\leq$ 365, days > 365
Hazard management:		
rIPC	Inspection point completion ratio	Numerical-%
rHD	Hidden danger ratio	Numerical-%
Hidden dangers_Rec	Rectification of hidden dangers	Numerical
Hidden dangers_ h Level	The highest level of hidden dangers	Nominal-Level I, Level II, Level III
Unit fire data maintenance:		
RegulatoryUnitsTyp	Types of Regulatory Units	Nominal-Yes/No
FCRL	Fire control room location	Nominal-Yes/No
UPC	Unit property category	Nominal-Yes/No
BgLayout	Building layout	Nominal-Yes/No
NO_EvacuationSairs	Number of evacuation stairs	Numerical
NO_SafeExits	Number of safe exits	Numerical
FFAEEPlans	Fire fighting and emergency evacuation plans	Nominal-Yes/No
FS_Sys	Fire safety system	Nominal-Yes/No
FS_Res	Fire safety responsible person	Nominal-Yes/No
FS_Man	Fire safety manager	Nominal-Yes/No
FS_Lia	Fire safety liaison	Nominal-Yes/No
FSupervisor	Fire supervisor	Nominal-Yes/No

References

1. Zheng, W. Fire Safety Assessment of China's Twelfth National Games Stadiums. *Procedia Eng.* **2014**, *71*, 95–100. [CrossRef]
2. Hamed, T.; Dara, R.; Kremer, S.C. Network intrusion detection system based on recursive feature addition and bigram technique. *Comput. Secur.* **2018**, *73*, 137–155. [CrossRef]
3. Latah, M.; Toker, L. Towards an efficient anomaly-based intrusion detection for software-defined networks. *IET Netw.* **2018**, *7*, 453–459. [CrossRef]
4. Zou, Q.; Zhang, T.; Liu, W. A fire risk assessment method based on the combination of quantified safety checklist and structure entropy weight for shopping malls. *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.* **2021**, *235*, 610–626. [CrossRef]
5. Choi, J.-H.; Lee, S.-W.; Hong, W.-H. A development of fire risk map and risk assessment model for urban residential areas by raking fire causes. *J. Archit. Inst. Korea Plan. Des.* **2013**, *29*, 271–278.
6. Liu, F.; Zhao, S.; Weng, M.; Liu, Y. Fire risk assessment for large-scale commercial buildings based on structure entropy weight method. *Saf. Sci.* **2017**, *94*, 26–40. [CrossRef]
7. Wang, S.-H.; Wang, W.-C.; Wang, K.-C.; Shih, S.-Y. Applying building information modeling to support fire safety management. *Autom. Constr.* **2015**, *59*, 158–167. [CrossRef]
8. Cheng, X.-Q.; Jin, X.L.; Wang, Y.; Guo, J.; Zhang, T.; Li, G. Survey on big data system and analytic technology. *J. Softw.* **2014**, *25*, 1889–1908.
9. Lo, S.M.; Liu, M.; Zhang, P.H.; Yuen, K.K.R. An Artificial Neural-network Based Predictive Model for Pre-evacuation Human Response in Domestic Building Fire. *Fire Technol.* **2008**, *45*, 431–449. [CrossRef]
10. Madaio, M.; Chen, S.-T.; Haimson, O.L.; Zhang, W.; Cheng, X.; Hinds-Aldrich, M.; Chau, D.H.; Dilkina, B. Firebird: Predicting fire risk and prioritizing fire inspections in Atlanta. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New Orleans, LA, USA, 13–17 August 2016; pp. 185–194.

11. Kim, D.H. A study on the development of a fire site risk prediction model based on initial information using big data analysis. *J. Soc. Disaster Inf.* **2021**, *17*, 245–253.
12. Liu, Z.-G.; Li, X.-Y.; Jomaas, G. Identifying Community Fire Hazards from Citizen Communication by Applying Transfer Learning and Machine Learning Techniques. *Fire Technol.* **2020**, *57*, 2809–2838. [CrossRef]
13. Surya, L. Risk Analysis Model That Uses Machine Learning to Predict the Likelihood of a Fire Occurring at A Given Property. *Int. J. Creat. Res. Thoughts (IJCRT) ISSN* **2017**, *5*, 2320–2882.
14. Anderson-Bell, J.; Schillaci, C.; Lipani, A. Predicting non-residential building fire risk using geospatial information and convolutional neural networks. *Remote Sens. Appl. Soc. Environ.* **2021**, *21*, 100470. [CrossRef]
15. Sayad, Y.O.; Mousannif, H.; Al Moatassime, H. Predictive modeling of wildfires: A new dataset and machine learning approach. *Fire Saf. J.* **2019**, *104*, 130–146. [CrossRef]
16. Xie, H.; Weerasekara, N.N.; Issa, R.R.A. Improved System for Modeling and Simulating Stadium Evacuation Plans. *J. Comput. Civ. Eng.* **2017**, *31*, 04016065. [CrossRef]
17. Darst, B.F.; Malecki, K.C.; Engelman, C.D. Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. *BMC Genet.* **2018**, *19*, 1–6. [CrossRef]
18. Zhu, X.; Khosravi, M.; Vaferi, B.; Amar, M.N.; Ghriga, M.A.; Mohammed, A.H. Application of machine learning methods for estimating and comparing the sulfur dioxide absorption capacity of a variety of deep eutectic solvents. *J. Clean. Prod.* **2022**, *363*, 132465. [CrossRef]
19. Zhu, H.; You, X.; Liu, S. Multiple Ant Colony Optimization Based on Pearson Correlation Coefficient. *IEEE Access* **2019**, *7*, 61628–61638. [CrossRef]
20. Chicco, D.; Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genom.* **2020**, *21*, 1–13. [CrossRef]
21. Refaeilzadeh, P.; Tang, L.; Liu, H. *Cross-Validation. Encyclopedia of Database Systems*; Springer: New York, NY, USA, 2009; Volume 5, pp. 532–538.
22. Haixiang, G.; Yijing, L.; Shang, J.; Mingyun, G.; Yuanyue, H.; Bing, G. Learning from class-imbalanced data: Review of methods and applications. *Expert Syst. Appl.* **2017**, *73*, 220–239. [CrossRef]
23. Poh, C.Q.; Ubeynarayana, C.U.; Goh, Y.M. Safety leading indicators for construction sites: A machine learning approach. *Autom. Constr.* **2018**, *93*, 375–386. [CrossRef]
24. Guan, F.; Shi, J.; Ma, X.; Cui, W.; Wu, J. A method of false alarm recognition based on k-nearest neighbor. In Proceedings of the 2017 International Conference on Dependable Systems and Their Applications (DSA), Beijing, China, 31 October–2 November 2017.
25. Gholizadeh, P.; Esmaeili, B.; Memarian, B. Evaluating the Performance of Machine Learning Algorithms on Construction Accidents: An Application of ROC Curves. In *Construction Research Congress 2018*; ASCE: Washington, DC, USA, 2018. [CrossRef]
26. Dang, T.T.; Cheng, Y.; Mann, J.; Hawick, K.; Li, Q. Fire risk prediction using multi-source data: A case study in humberside area. In Proceedings of the 2019 25th International Conference on Automation and Computing (ICAC), Lancaster, UK, 5–7 September 2019; pp. 1–6.
27. Zhu, R.; Hu, X.; Hou, J.; Li, X. Application of machine learning techniques for predicting the consequences of construction accidents in China. *Process Saf. Environ. Prot.* **2020**, *145*, 293–302. [CrossRef]
28. Pirklbauer, K.; Findling, R.D. Storm Operation Prediction: Modeling the Occurrence of Storm Operations for Fire Stations. In Proceedings of the 2021 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events, Kassel, Germany, 22–26 March 2021; pp. 123–128.
29. Wang, Q.; Zhang, J.; Guo, B.; Hao, Z.; Zhou, Y.; Sun, J.; Yu, Z.; Zheng, Y. CityGuard: Citywide fire risk forecasting using a machine learning approach. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2019**, *3*, 1–21. [CrossRef]
30. Zhang, Y.; Geng, P.; Sivaparthipan, C.; Muthu, B.A. Big data and artificial intelligence based early risk warning system of fire hazard for smart cities. *Sustain. Energy Technol. Assess.* **2021**, *45*, 100986. [CrossRef]
31. Chang, J.; Yoon, J.; Lee, G. Machine Learning Techniques in Structural Fire Risk Prediction. *Int. J. Softw. Eng. Its Appl.* **2020**, *14*, 17–26. [CrossRef]