Heliyon 9 (2023) e19092

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Comparative study of multiple machine learning algorithms for risk level prediction in goaf

Bin Zhang **, Shaohua Hu, Moxiao Li *

School of Safety Science and Emergency Management, Wuhan University of Technology, Wuhan, Hubei, 430070, China

ARTICLE INFO

Keywords:
Mine safety
Machine learning
Data processing
Risk prediction
Data classification

ABSTRACT

With the acceleration of the mining process, the goaf has become one of the main sources of danger in underground mines, seriously threatening the safe production of mines. To make an accurate prediction of the risk level of the goaf quickly, this paper optimizes the features of the goaf by correlation analysis and feature importance and constructs a combination of feature parameters for the risk level prediction of the goaf to solve the problem of redundancy of evaluation indexes. Multiple machine learning algorithms are applied to 121 sets of goaf data respectively, and the optimal algorithm and the best combination of feature parameters are obtained by evaluating the mining area with multiple indicators such as accuracy and kappa coefficient. The best combination of features parameters are ground-water, goaf layout, volume of goaf, goaf volume, span-height ratio, and mining disturbance, and the optimal algorithm is Extra Tree (ET), which needles the goaf risk level prediction problem with the accuracy of 94%. This model can be used to solve the problem of how to quickly and accurately predict the risk level of the goaf.

1. Introduction

In recent years, with the acceleration of the mining process worldwide, the number of goaves has gradually increased, and the goaf has become one of the main sources of danger in underground mines, caused serious casualties and property damage, seriously threatening the safe production of mines [1]. Untreated goaf will cause surface collapse, a large area of rock block falling, or the formation of goaf ponding, which will lead to water seepage of the working face, causing casualties and property losses [2,3]. Currently, there are several key points in research on goaf areas: firstly, spontaneous combustion in goaf areas [4]; secondly, the danger of methane explosions in goaf areas [5,6]; thirdly, the issue of goaf collapse. Therefore, it is of great significance to carry out stability evaluation and risk prediction of goaf for mine safety production. This study primarily focuses on the third point, utilizing methods such as data mining and machine learning to rapidly and accurately predict the stability level of goaf areas, thus contributing to the safety of underground mining operations.

For the evaluation of the stability of goaf, more methods have been proposed by domestic and foreign scholars. HU et al. and Yan et al. [7,8]proposed a Bayesian discriminant analysis method to identify the hazard of complex goaf in mines. The method identifies the stability of goaf by selecting nine factors affecting the risk stability of goaf as discriminant indicators and establishing a discriminant analysis model. Dong et al. [9]based on uncertainty metric evaluation (WME) and hierarchical analysis (AHP) theory,

* Corresponding author.
** Corresponding author.
E-mail addresses: bin1075@whut.edu.cn (B. Zhang), lmx@whut.edu.cn (M. Li).

https://doi.org/10.1016/j.heliyon.2023.e19092
Received 24 February 2023; Received in revised form 9 August 2023; Accepted 10 August 2023
Available online 11 August 2023

used multiple indicators to conduct a comprehensive risk evaluation of goaf. Ma et al. [10]used a mechanical method combined with the Voronoi diagram method to establish a dynamic analysis model for the stability of the coal pillar-roof system. Sun et al. [11] combined theoretical analysis, geological and geophysical investigation, found four anomalous zones, and analyzed that mining and hydrology are the two important factors affecting goaf. However, the above methods did not consider the problem of too many data features in the extraction zone and the duplication of the information contained, and there is usually some redundancy in the selection of indicators and the use of data.

To avoid the problem of selecting complex indicators, some scholars used numerical simulation to achieve the stability judgment of the mining area [12]. Ao et al. [13]proposed a 3D geoengineering model to depict intricate geological formations and formulated a corresponding 3D stability analysis model grounded in geotechnical principles. This enabled the assessment of grouting reinforcement's impact and mining area stability. Gao et al. [14]used the synthetic rock mass (SRM) model to obtain rock properties and used UDEC triangulation to simulate roadway damage. Shi et al. [15] developed a numerical computation model to analyze various surface deformation indicators across distinct mining sequences of multiple coal seams. This investigation unveiled the temporal impact of mining operations on surface deformation patterns within the subsidence zone, facilitating the determination of the most favorable mining sequence. However, the numerical simulation process needs to simplify the boundary conditions and other phenomena, which cannot truly restore the complex geological structure of the mining area and other engineering parameters, and the simulation method is usually based on a certain theory or framework, once the theory or framework has problems, it will make the simulation results differ greatly from the actual, so that the accuracy of the simulation results cannot be guaranteed. In conclusion, to address the issue of excessive redundancy in evaluation indicators for goaf, this study has decided to employ the Pearson correlation coefficient to calculate the correlation coefficients between indicators, following the standards set by scientific research. Additionally, by integrating machine learning techniques, the study aims to determine the importance of evaluation indicators. The combination of these two methods will yield the optimal combination of evaluation indicators suitable for predicting the stability of goaf.

With the continuous development of artificial intelligence and machine learning in recent years, relevant intelligent algorithms have been gradually used in underground mine engineering fields. Zhou et al. [16]used ten supervised learning methods to predict the rockburst intensity level, and compared with the traditional rockburst intensity level prediction methods, the machine learning-related methods have the advantages of high accuracy and high speed; Qi et al. [17]used five machine learning algorithms to The stability of the empty field hanging gang was predicted and the hyperparameters of the five machine learning algorithms were optimized by the Firefly algorithm, and good results were achieved. Jarosław Brodny et al. [18] employed a neuro-fuzzy model for near-term methane prediction in coal mines. This model integrated actual ventilation parameter measurements and harnessed the Methane Hazard Index derived from the neural fuzzy model, they evaluated and forecasted the current level of methane hazard. This research approach provides a novel and innovative method for assessing methane hazards. Prasanjit Dey et al. [19]combined Internet of Things (IoT) sensor devices with machine learning methods, employing a hybrid CNN-LSTM network to predict the miner's health quality index and CH4 gas concentration in coal mines. Their approach yielded promising results, showcasing its effectiveness. Arif Hussain Soomro et al. [20]utilized wireless sensor networks (WSN) and artificial neural networks (ANN) to predict the outburst of gas, ensuring the safety of miners.

As the monitoring data of the goaf often present multidimensional and nonlinear characteristics, a large amount of information is contained in different dimensional data, and intelligent algorithms have certain advantages for mining the implicit information in these data. In recent years, gradually some scholars have combined machine learning with the problem of goaf, Wang et al. [21]used wavelet support vector machine to predict the residual subsidence of abandoned goaf; Luan et al. [2]used the support vector machine approach for stability prediction of goaf; Qin et al. [22]used the improved TrAdaBoost algorithm based on the migration learning theory that predicts the stability level of the goaf. However, the problem of a large number of dimensions and strong correlation of the data in goaf remains unresolved, and the use of machine learning methods often requires a large amount of data as support [23], and the amount of sample data of the goaf used in previous studies is small and not sufficient to reveal the universal laws.

Based on this, this paper combines correlation analysis and feature importance analysis to optimize the feature parameters of the goaf, and constructs a combination of feature parameters for the evaluation of the stability of the goaf to solve the problem of redundancy of indicators of the goaf; according to the data characteristics of high-dimensional small samples of the goaf, various machine learning algorithms applicable to this problem are applied to 121 sets of goaf data, and the goaf is evaluated by accuracy, kappa coefficient, precision, recall and F-measure. In this paper, we obtained the optimal algorithm and the best combination of feature parameters for the evaluation of the stability of the goaf and formed a model that can quickly and accurately predict the risk level of the goaf. This model is applicable for the rapid and accurate assessment and prediction of goaf stability in mines, making a valuable contribution to the safe extraction of underground minerals.

2. Method

2.1. Algorithm selection

Integration algorithms, being a prevalent machine learning technique, exhibit significant advantages across a multitude of domains. There are three types of integration algorithms, including bagging [24], boosting [25], and stacking [26].

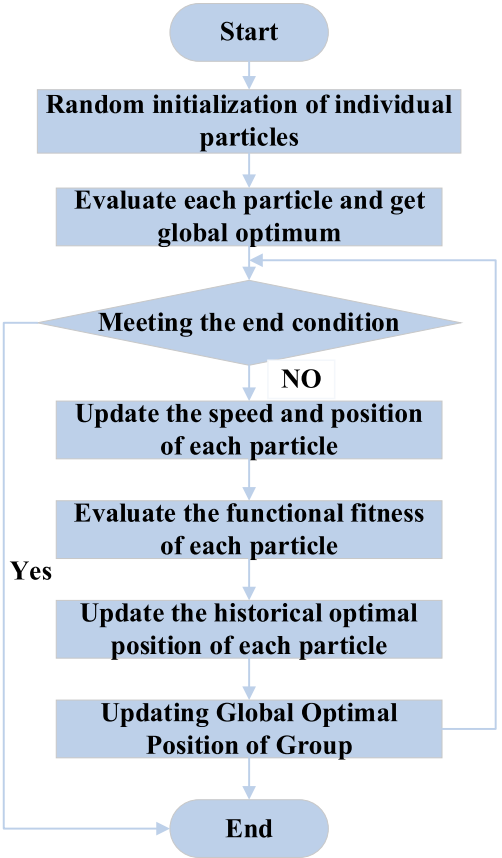
The main idea of bagging is to build multiple independent base evaluators, and then combine the results of each evaluator to get the final result, because each base evaluator does not correspond to the same training set, so multiple base evaluators are independent of each other; The boosting algorithm does not change the training set in each round, but continuously adjusts the weight values according to the error rate, so the base evaluators are interrelated, and it iterates through multiple base evaluators to arrive at the final

result; Stacking is a combination of different trained models, and each model is voted independently to reach the conclusion, which is not discussed in this paper because it is less interpretable and not commonly used in practice.

In this paper, we select typical algorithms for machine learning: Support Vector Machine (SVM) [27], Naive Bayes (NB) [28], K-nearest neighbor (KNN), Gradient Boosting Tree (GBDT) [29], Decision Tree (DT) [30], Multilayer Perceptron (MLP) and integrated algorithms: Random Forest (RF) [31,32], Extra Trees (ET) [33], Extreme Gradient Boosting algorithm (XGBoost) [34], AdaBoost (ADA) [35,36]. The reasons for adopting the above-mentioned methods in this paper are mainly threefold: (a) The selected algorithms in this study are widely applied in the field of engineering and demonstrate relatively good predictive performance. (b) Predicting the stability level of goaf areas involves a multidimensional, nonlinear computational problem. The machine learning algorithms mentioned above can utilize multiple input variables and learn nonlinear relationships, making them suitable for analyzing goaf stability issues. (c) Some machine learning algorithms are considered top-tier data mining algorithms. Comprehensive introductions and utilization methods of these algorithms can be located within the respective references cited in this paper.

2.2. PSO algorithm

Particle swarm optimization (PSO) is one of the most widely used optimization algorithms, applicable in various fields such as electrical and electronic engineering, automation control systems, communication theory, operations research, mechanical engineering, fuel and energy, medicine, chemistry, and biology [37]. Its concept is derived from the study of bird feeding behavior, where individuals in a bird group collaborate and share information to find the optimal solution. The specific algorithm flow and pseudo-code are shown in Fig. 1. The advantages of the PSO algorithm can be summarized as follows: (1) It exhibits good robustness and can be easily adapted to different application environments with minor modifications. (2) It possesses strong distributed capabilities, as it is essentially a population-based evolutionary algorithm, making it suitable for parallel computing implementations. (3) It can quickly converge to the optimal value. (4) It is easily combinable with other algorithms to enhance performance [38,39].



PSO algorithm pseudocode

```
procedure PSO
  for each particle I
    Initialize velocity Vi and position Xi for particle i
    Evaluate particle i and set pBesti = Xi
  end for
  gBest = min{pBesti}
  while not stop
    for i=1 to N
      Update the velocity and position of particle i
      Evaluate particle i
      if fit (Xi)<fit(pBesti)
        pBesti = Xi;
      if fit(pBesti)< fit(gBest)
        gBest = pBesti;
    end for
  end while
end procedure
```

Fig. 1. PSO algorithm flow and pseudocode.

3. Index analysis and data preprocessing

3.1. Index analysis

The factors affecting the stability of the goaf are mainly rock mechanics factors, environmental factors of the goaf. Among the quantitative indicators are: rock compressive strength X3, goaf volume X6, the exposed area of goaf roof X7, buried depth X8, and span-height ratio X9, and the qualitative indicators are: rock structure X1, geological structure X2, ground-water X4, goaf layout X5, mining disturbance X10, and condition of adjacent goaf X11. The goaf's structure in three categories, as shown in Fig. 2. And the stability level of goaf is divided into four categories, as shown in Table 1.

3.2. Data preprocessing

In this paper, 121 sets of data on the goaf were collected by reviewing the literature [40–42]. Among them, the risk level I is 25 groups, II is 40 groups, III is 39 groups, and IV is 17 groups.

Because of the large number of features of the collected goaf data and the presence of missing values of individual features, the data need to be pre-processed. For some of the data with missing values, by comparing the effects of filling methods using the mean, median, quantile, plurality, and random values of normal data, the median interpolation method was finally used to interpolate the ground-water of 34 groups of data present in the dataset, and 17 groups of goaf arrangements to minimize the impact of their missing data on the overall sample.

In this paper, we use the ggplot function in R language [43,44] to visualize the distribution and correlation of the data, as shown in Fig. 3 (a). The upper half indicates the correlation coefficients between the two variables, and the middle part indicates the probability distribution of indicators corresponding to the four levels of risk. The bottom half represents the distribution of data between the two variables, which shows that the dispersion of the data is high and the quality of the data set is good.

In the field of machine learning, unbalanced datasets can lead to overfitting, and the risk class datasets collected in this paper are unbalanced, including 25 sets of data for hazard class I, 40 sets of data for hazard class II, 39 sets of data for hazard class III, and 17 sets of data for hazard class IV. The SMOTE algorithm [45,46]uses a linear interpolation method between two minority class samples synthesizing new samples, thus effectively alleviating the overfitting problem caused by the imbalance of the data set. In this paper, the SMOTE algorithm is used to oversample the dataset, and the processed dataset is shown in Fig. 3 (b), with the data at four levels of 40, 40, 39, and 40.

Because of the large differences in the values and magnitudes of each feature of the collected goaf data, this paper standardizes the data of each feature to reduce the influence of factors such as large differences in feature values and different magnitudes on the

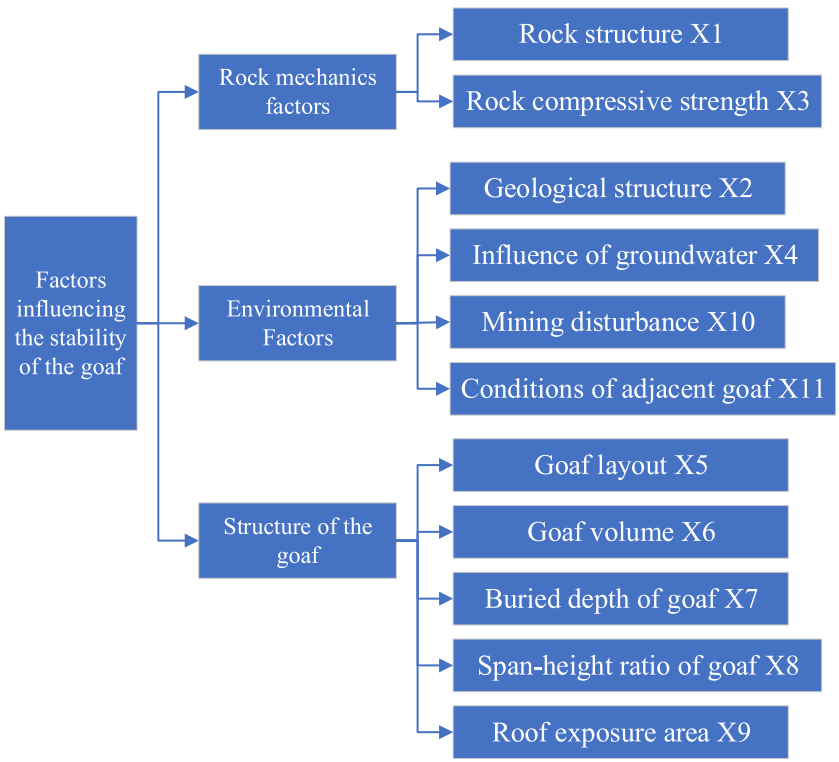


Fig. 2. Indicators affecting the risk level of the goaf.

Table 1
Stability class of goaf.

Stability level	Stability status or risk of instability
I	The goaf very good stability and no risk of instability
II	The goaf good stability and low risk of instability
III	The goaf is less stable and the risk of instability is greater
IV	Stability of the goaf is extremely poor and the risk of instability is very high

evaluation model. The RobustScaler method is used because of the existence of outliers in the actual collected mining area data set, and its standardization is shown in equation (1), where median is the median of the sample data and IQR is interquartile of the sample data. This method is used to maximize the retention of outliers in the dataset, thus enhancing the robustness of the model. In machine learning, the commonly used data partitioning methods are 70% (training set)/30% (testing set) or 80% (training set)/20% (testing set). However, due to the limited size of the goaf dataset used in this study, which consists of only 121 instances, the number of instances available for testing is relatively small. Using only 20% of the data as the testing set would result in a test set of only 32 instances, which would introduce greater randomness in the results. Therefore, we have chosen to allocate 30% of the data as the testing set to ensure more reliable and robust testing results. Then to improve the generalization of the model, the dataset is randomly disrupted using the function of shuffle.

$$v_i^{'} = \frac{v_i - \text{median}}{IQR} \tag{1}$$

3.3. Hyperparameter tuning

Parameter tuning is an important task in machine learning. In this paper, the PSO is used to adjust the hyperparameters of the following algorithm, and the specific tuning results are shown in Table 2.

4. Feature combination selection

4.1. Analysis of relationship

The existence of a correlation between each feature can hurt make the accuracy of machine learning decrease. In this paper, the correlation coefficient between each feature is calculated using the Pearson correlation coefficient method [47,48], which is calculated as in equation (2), where $\mathrm{cov}(X, Y)$ represents the covariance between X and Y variables, and σ is the standard deviation of the variables. In this paper, the correlation coefficients between each feature are visualized in Fig. 2, and the magnitude of the absolute value of the value represents the strength of the correlation between the features, with a negative sign representing the negative correlation between the features and the opposite positive correlation, where * is an average correlation, ** is a strong correlation, and *** is a significant correlation.

The correlations between the two feature variables were counted by number and the weights of ***、** and * were set to 0.3, 0.2, and 0.1, respectively, to derive the correlations between each feature and the remaining features. From Fig. 2 and Table 3, it can be seen that the correlations among the 11 features are strong, so it is necessary to reduce the features. After analysis, X3, X5, X,6 and X7 have relatively low correlations with other features, so they can be regarded as independent features for selection.

$$\rho_{X,Y} = \frac{\mathrm{cov}\ (X, Y)}{\sigma_X \sigma_Y} = \frac{E((X - \mu_X)(Y - \mu_Y))}{\sigma_X \sigma_Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E\left(X^2\right) - E^2(X)}\sqrt{E\left(Y^2\right) - E^2(Y)}} \tag{2}$$

4.2. Feature importance analysis

In this paper, the following six algorithms: Random Forest (RF), Xgboost (XGB), ExtraTrees (ET), Adaboost (ADA), Gradient Boosting Decision Trees (GBDT), and Decision Trees (DT) are used to rank the importance of features affecting the goaf. The features affecting the stability of the goaf are ranked in importance, and the ranking results are shown in Fig. 3 (a) below.

Fig. 4(a–f) shows the ranking of the importance of the features by different algorithms for the stability problem of the goaf. The gray shaded part in the figure represents the threshold value of the importance of 0.10. When the importance of a feature exceeds 0.1 in the corresponding algorithm, the corresponding position in Table 4 is marked with a tick. From Fig. 4(a–f) and Tables 4 and it can be concluded that the importance of features X4, X5, X6, X7, X9, and X10 is relatively high, while the importance of X1, X2, X8, X11, etc. is relatively low.

Combining the above feature correlation and importance analysis with the actual engineering situation, nine sets of feature combinations applicable to the stability risk level prediction of the goaf are derived, as shown in Table 5.

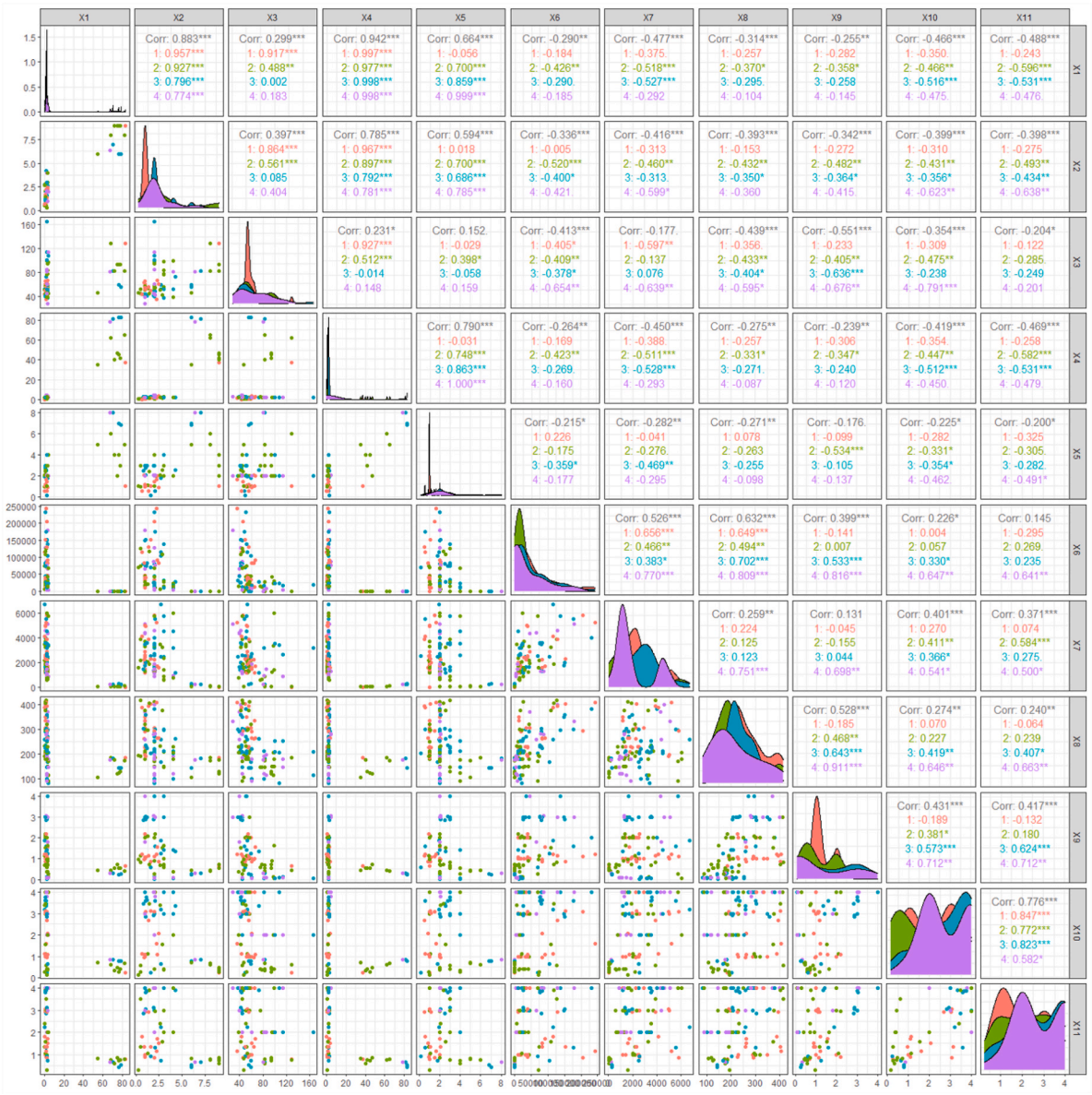


Fig. 3. Distribution and correlation of data sets (a illustrates the visualization of goaf data, while b presents a comparison of the proportions of the four goaf grades before and after preprocessing).

Table 2
Results of hyperparameter regulation by different algorithms.

Algorithm	Tuning parameters	Explanation	Settings
RF	n_estimators	The number of trees in the forest	186
	max_depth	The maximum depth of the tree	auto
	min_samples_leaf	The minimum number of samples required to split an internal node	2
	max_samples_split	The minimum number of samples required to be at a leaf node	1
GNB	prior	the prior probability of the class	None
	var_smoothing		0.000000001
GBDT	loss	Loss function to be optimized	log_loss
	n_estimators	Number of advancement phases to be executed	175
	learning_rate	Learning Rate	0.1
DT	criterion	Feature selection criteria	Gini
	splitter	Characteristics classification criteria	best
	min_samples_leaf	The minimum number of samples required to be at a leaf node	1
	min_samples_split	The minimum number of samples required to split an internal node	2
SVM	C	Penalty Factor	1.0
	kernel	Specifies the kernel type to be used in the algorithm	RBF
KNN	n_neighbors	The value of k in KNN	1
	weights	The weights of the nearest neighbor samples for each sample	uniform
	algorithm	The algorithm used for the restricted radius nearest neighbor method	auto
MLP	hidden_layer_sizes	The ith element represents the number of neurons in the ith hidden layer	286
	activation	Activation function for the hidden layer	relu
XGB	booster	Iterative application of the model	gbtree
	learning_rate	Learning Rate	0.05
	n_estimators	The number of weak learners	167
ADA	n_estimators	The number of weak learners	154
	learning_rate	the contribution of the weak learners in the final combination	0.8
	base_estimator	Base classifier	estimatorCart
ET	n_estimators	The number of trees in the forest	41
	criterion	The function to measure the quality of a split	Gini
	max_depth	The maximum depth of the tree	14
	min_samples_leaf	The minimum number of samples required to split an internal node	1

Table 3
Correlation coefficient score.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
***	8	10	6	6	3	5	6	5	6	7	6
**	2	0	0	3	2	2	2	5	2	1	1
*	0	0	2	1	3	2	0	0	0	2	2
scores	2.8	3	2	2.5	1.6	2.1	2.2	2.5	2.2	2.5	2.2

5. Model construction and evaluation indexes

5.1. Model construction

The technical route for the construction of the risk level prediction model for the extraction area is shown in Fig. 5, which includes the following steps.

(1) The 121 sets of risk level data samples collected from the mining area were pre-processed by applying SMOTE and RobustScaler methods to construct a database applicable to the risk level prediction model of the mining area.
(2) A combination of correlation analysis and feature importance ranking was used to construct a combination of features suitable for this study.
(3) The selected feature combinations and the selected machine learning algorithms are combined to perform ten-fold cross-validation, hyperparameter optimization, etc.
(4) A multi-indicator is used to evaluate the model and derive the most suitable feature combination and machine learning algorithm for the risk level prediction model of the goaf.

5.2. Model evaluation methods

For the prediction task of subsidence risk levels in goaf areas, accuracy, kappa coefficient, precision, recall, and F1 score are employed as the evaluation criteria for the proposed model in this paper. These metrics are computed through the confusion matrix as shown in Fig. 6.

Accuracy is a measure of whether a multiclassification problem is classified accurately, is calculated as in equation (3), where n is

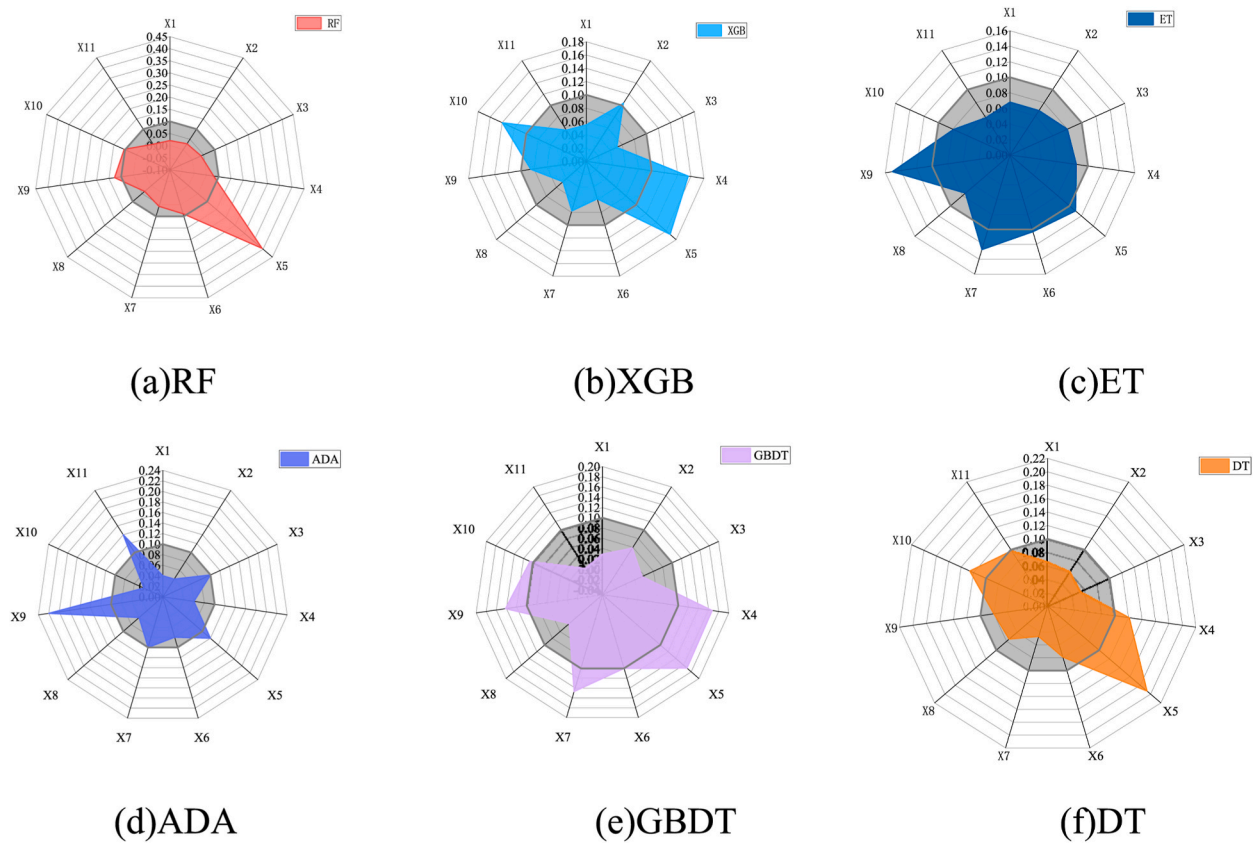


Fig. 4. Feature importance ranking chart (a to f represent the importance of calculating the influence of goaf indicators using the RF, XGB, ET, ADA, GBDT, and DT algorithms).

Table 4
Sorting the importance of each feature (corresponding to the use of $\sqrt{}$ tags with importance exceeding 0.1).

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
RF					✓				✓	✓	
XGB		✓		✓	✓					✓	
ET					✓	✓	✓		✓		
ADA			✓		✓		✓		✓		✓
GBDT				✓	✓	✓	✓		✓	✓	
DT			✓	✓	✓					✓	

Table 5
Combination of evaluation features selected in this article.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
A	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
B		✓	✓	✓	✓	✓	✓		✓	✓	✓
C			✓	✓	✓	✓	✓		✓	✓	✓
D			✓	✓	✓	✓	✓		✓	✓	
E				✓	✓	✓	✓		✓	✓	
F			✓		✓	✓	✓		✓	✓	
G			✓			✓	✓		✓	✓	
H			✓		✓	✓	✓		✓	✓	
I			✓		✓	✓	✓			✓	
J			✓		✓	✓	✓		✓		

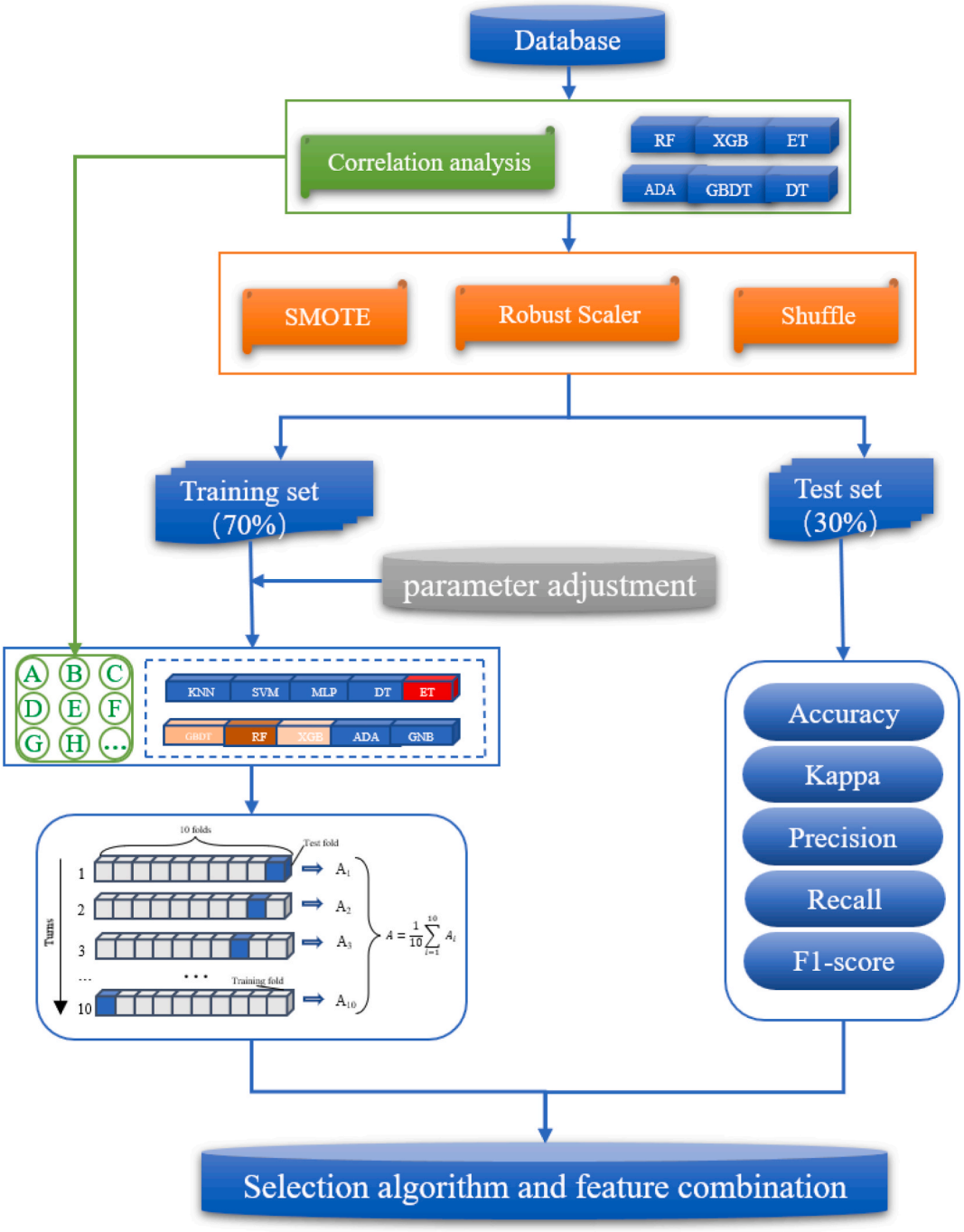


Fig. 5. Technical roadmap.

the total number of samples, C is the number of sample categories, and x_{ii} is the number of samples on the diagonal of the confusion matrix.

$$\text{Accuracy} = \left(\frac{1}{n}\sum_{i=1}^{C} x_{ii}\right) \tag{3}$$

The kappa coefficient is an index to evaluate the merits of the classification model. The kappa coefficient is taken as 0.0–0.20 for slight agreement, 0.21–0.40 for fair agreement, 0.41–0.60 for moderate agreement, and 0.61–0.80 for substantial agreement, 0.81–1 almost perfect, calculated as in equation (4), where n is the total number of samples, C is the number of sample categories, xii is the

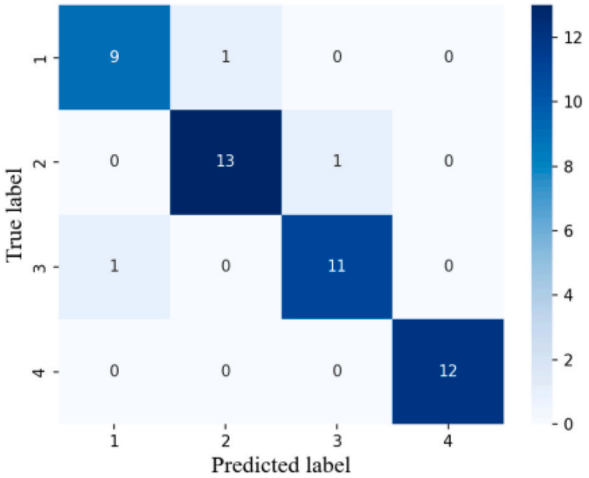


Fig. 6. Confusion matrix of partial algorithm.

number of samples on the diagonal of the confusion matrix, $x_{i.}$ and $x_{.\ i}$ are the number of samples in the rows and columns of the confusion matrix, respectively.

$$\text{Kappa} = \frac{n\sum i = 1^{C} x_{ii} - \sum i = 1^{C}(x_{i.} \cdot x_{.i})}{n^2 - \sum i = 1^{C}(x_{i.} \cdot x_{.i})} \tag{4}$$

Recall, Precision, and F1 values are important indicators of the classification problem, which are calculated as equation (5), equation (6), and equation (7), respectively.

$$\text{Recall}_i = \left(\frac{x_{ii}}{x_{+C}}\right) \times 100\% = \left(\frac{x_{ii}}{\sum_{i=1}^{C} x_{iC}}\right) \times 100\% \tag{5}$$

$$\text{Precision}_i = \left(\frac{x_{ii}}{x_{+C}}\right) \times 100\% = \left(\frac{x_{ii}}{\sum_{i=1}^{C} x_{iC}}\right) \times 100\% \tag{6}$$

$$F - \text{measure} = \frac{2\ \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{7}$$

6. Results and discussion

6.1. Results analysis

The accuracy of different algorithms and their corresponding feature combinations on the test set is obtained by combining them as shown in Table 6. For a clearer and more intuitive representation of its accuracy and the size of the kappa coefficient, the results are drawn into a three-dimensional graph as shown in Fig. 7(a–b), where the x y axis represents the feature combination and the algorithm used, and the z axis represents the size of the corresponding value.

From Fig. 7 and Tables 6 and it can be seen that the best-performing model is the ET, the accuracy and kappa coefficient of this model are significantly better than other algorithms, with the highest accuracy of 94% and the average accuracy of 88.9%; the highest kappa coefficient reaches 0.916 and the average kappa coefficient reaches 0.8473, and the accuracy of this algorithm reaches 88% and above for all the feature groups except for the group I feature In addition to the 81% accuracy in Group I, the algorithm achieves 88% accuracy and above in the rest of the feature groups. The next model is the RF, with the highest accuracy of 88%, the average accuracy of 84%, the highest kappa coefficient of 0.834, and the average kappa coefficient of 0.79. The algorithm is slightly less sensitive to feature selection than the ET, and achieves 77% on the group I features. In addition, the XGB and KNN models also performed well, with the XGB model achieving a maximum accuracy of 88% with a kappa coefficient of 0.833 and an average accuracy of 83.56% with an average kappa coefficient of 0.781. The KNN model achieved a maximum accuracy of 88% with a kappa coefficient of 0.834, while the average accuracy was 83.7% with an average kappa coefficient of The average kappa coefficient was 0.784.

The GNB and ADA algorithms, on the other hand, performed poorly on this dataset, with the highest accuracy rates of only 58% and 60%, and the corresponding kappa coefficients of 0.45 and 0.464, respectively.

Table 6
Performance of ten algorithms in ten evaluation index combinations.

Feature combination	Evaluating indicator	Algorithm										Mean value
		KNN	SVM	MLP	DT	ET	GBDT	RF	XGB	ADA	GNB	
A	Accuracy	0.88	0.73	0.85	0.83	0.92	0.81	0.85	0.88	0.83	0.58	0.816
	kappa	0.834	0.641	0.807	0.778	0.888	0.751	0.806	0.832	0.778	0.45	0.7565
B	Accuracy	0.85	0.71	0.83	0.79	0.9	0.83	0.88	0.83	0.92	0.6	0.814
	kappa	0.805	0.616	0.779	0.723	0.86	0.779	0.834	0.779	0.889	0.478	0.7542
C	Accuracy	0.88	0.77	0.79	0.79	0.88	0.81	0.83	0.83	0.83	0.65	0.806
	kappa	0.833	0.697	0.723	0.724	0.834	0.751	0.778	0.778	0.777	0.533	0.7428
D	Accuracy	0.88	0.67	0.75	0.79	0.9	0.81	0.85	0.81	0.88	0.54	0.788
	kappa	0.833	0.559	0.667	0.724	0.862	0.751	0.806	0.75	0.833	0.394	0.7179
E	Accuracy	0.83	0.75	0.81	0.81	**0.92**	0.83	0.85	0.9	0.9	0.62	**0.828**
	kappa	0.778	0.669	0.751	0.751	**0.888**	0.779	0.806	0.861	0.86	0.422	**0.7565**
F	Accuracy	0.85	0.71	0.77	0.79	0.88	0.81	0.83	0.81	0.9	0.56	0.791
	kappa	0.806	0.616	0.696	0.721	0.834	0.751	0.778	0.749	0.861	0.422	0.7234
G	Accuracy	0.85	0.67	0.71	0.79	0.9	0.88	0.85	0.88	0.88	0.52	0.793
	kappa	0.806	0.558	0.606	0.723	0.86	0.834	0.805	0.833	0.832	0.364	0.7221
H	Accuracy	0.85	0.75	0.79	0.81	**0.94**	0.81	0.85	0.88	0.88	0.5	0.806
	kappa	0.807	0.669	0.723	0.75	**0.916**	0.75	0.805	0.833	0.832	0.334	0.7419
I	Accuracy	0.83	0.69	0.77	0.75	0.81	0.77	0.85	0.77	0.81	0.44	0.749
	kappa	0.779	0.588	0.694	0.668	0.751	0.696	0.806	0.694	0.751	0.251	0.6678
J	Accuracy	0.81	0.71	0.75	0.79	0.88	0.85	0.85	0.83	0.85	0.56	0.788
	kappa	0.75	0.615	0.666	0.72	0.834	0.805	0.805	0.778	0.805	0.421	0.7199
	Mean accuracy	0.851	0.716	0.782	0.794	**0.893**	0.821	0.849	0.842	0.868	0.557	
	Mean kappa	0.8031	0.6228	0.7112	0.7282	**0.8527**	0.7647	0.8029	0.7887	0.8218	0.4069	

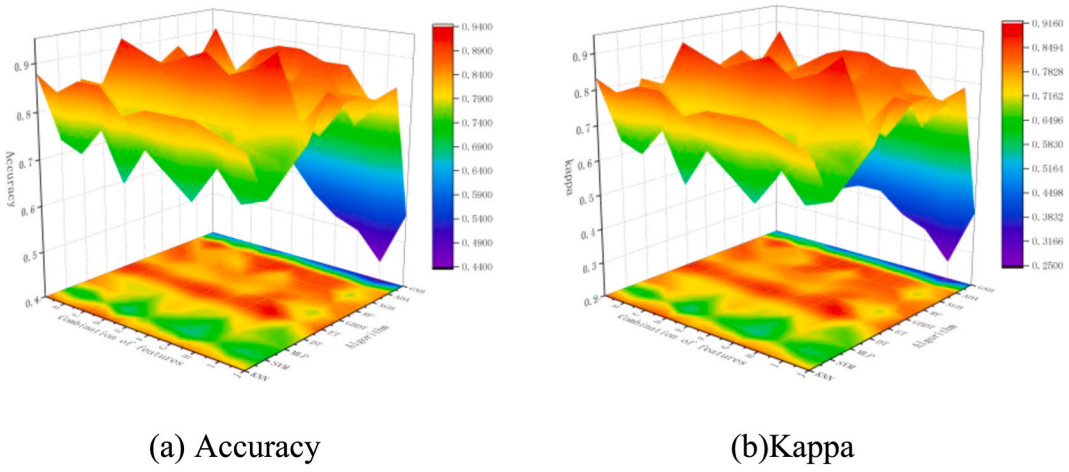


Fig. 7. Accuracy and kappa coefficient (X,Y axis represents the feature combination and the algorithm used, and the Z axis represents the size of the corresponding value. Figure a illustrates the accuracy values corresponding to different algorithms and feature combinations, b represents the kappa values corresponding to different algorithms and feature combinations).

In order to verify the generalizability of the algorithm, the training set is shuffled in this paper, and then the dataset is tested using k-fold cross-validation, which divides the entire dataset into k copies, so that k-1 copies are used as the training set, leaving one copy as the test set, and the accuracy of this test set is obtained, and the above work is repeated k times, and the accuracy of k times of validation is averaged as the final accuracy. The value of k is chosen as 10 recommended by T. T. Wong et al. [34], and the accuracy rates of different algorithm combinations and their corresponding feature combinations are calculated as shown in the following table.

As can be seen from Fig. 8 and Table 7, in the cross-validation results, The highest mean accuracy values are for ET, RF, and ADA, and when features C and E are chosen, the mean accuracy values increase to 77.1% and 76.12%, respectively. In conclusion, based on the results of the ten-fold cross-validation, it can be seen that the extreme random tree ET, RF, and ADA algorithms have high generalizability.

In summary, the models that performed well in this task are ET, RF, and ADA, which have one thing in common: they are all based on the tree model as the base classifier model. While models such as SVM that use a distance metric to measure the class to which they belong perform generally. GNB performs poorly because the collected data set has more noise and does not exactly match the normal distribution, while GNB performs better when the data is in a normal distribution and less well when there is more noise. The average value of accuracy and kappa coefficient of group E features is the highest among the feature combinations, and it also performs well in cross-validation.

The above analysis was integrated to make ten algorithms to predict the risk level of the four-level mining area on the E group features, respectively, using Precision, Recall, and F-measure values as the measurement criteria, and the calculation results are shown in Table 8. Among them, the accuracy ranges from I to IV are 0.58–0.90; 0.56–1; 0.61–1; 0.6–1, the recall ranges are 0.7–1; 0.36–0.93; 0.5–0.92; 0.75–1, and the F1 values range from 0.64–0.91; 0.43–0.93; 0.57–0.87; 0.67–1, respectively, as shown in Fig. 9(a–c). Among

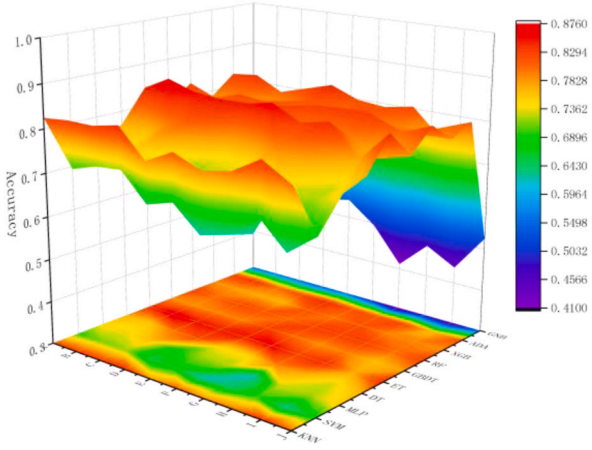


Fig. 8. Accuracy of 10-fold cross validation (X,Y axis represents the feature combination and the algorithm used, and the Z axis represents the size of the corresponding value).

Table 7
10 fold cross validation accuracy.

	KNN	SVM	MLP	DT	ET	GBDT	RF	XGB	ADA	GNB	Feature combination mean
A	0.825	0.694	0.744	0.769	0.844	0.787	0.831	0.787	0.831	0.537	0.7649
B	0.831	0.713	0.738	0.756	0.875	0.787	0.806	0.781	0.838	0.55	0.7675
C	0.831	0.719	0.762	0.8	0.856	0.794	0.825	0.775	0.806	0.544	0.7712
D	0.844	0.656	0.681	0.819	0.85	0.794	0.819	0.769	0.825	0.531	0.7588
E	0.781	0.675	0.719	0.744	0.863	0.831	0.806	0.818	0.844	0.544	0.7625
F	0.813	0.619	0.719	0.744	0.863	0.806	0.831	0.8	0.819	0.494	0.7508
G	0.788	0.644	0.619	0.794	0.831	0.787	0.831	0.819	0.813	0.412	0.7338
H	0.8	0.694	0.731	0.769	0.844	0.794	0.831	0.794	0.825	0.475	0.7557
I	0.8375	0.631	0.719	0.731	0.775	0.762	0.806	0.794	0.763	0.438	0.72565
J	0.8	0.681	0.75	0.769	0.844	0.781	0.832	0.788	0.813	0.525	0.7583
Algorithm mean	0.81505	0.6726	0.7182	0.7695	**0.8445**	0.7923	0.8218	0.7925	0.8177	0.505	

Table 8
Algorithm performance on Group E features.

Category	Evaluating indicator	KNN	SVM	MLP	DT	ET	GBDT	RF	XGB	ADA	GNB
I	Precision	0.71	0.58	0.69	0.71	0.83	0.71	0.69	0.75	0.90	0.60
	Recall	**1.00**	0.70	0.90	**1.00**	**1.00**	**1.00**	0.90	0.90	0.90	0.90
	F1-values	0.83	0.64	0.78	0.83	0.91	0.83	0.78	0.82	0.90	0.72
II	Precision	**1.00**	0.86	0.89	**1.00**	**1.00**	0.90	0.91	0.92	0.93	0.56
	Recall	0.64	0.43	0.57	0.64	0.86	0.64	0.71	0.79	0.93	0.36
	F1-values	0.78	0.57	0.70	0.78	0.92	0.75	0.80	0.85	0.93	0.43
III	Precision	**1.00**	0.61	0.79	0.90	0.91	0.79	0.91	0.91	0.91	0.67
	Recall	0.75	0.92	0.92	0.75	0.83	0.92	0.83	0.83	0.83	0.50
	F1-values	0.86	0.73	0.85	0.82	0.87	0.85	0.87	0.87	0.87	0.57
IV	Precision	0.75	**1.00**	**1.00**	0.80	0.92	**1.00**	0.92	0.92	0.92	0.60
	Recall	**1.00**	0.92	**1.00**	**1.00**	**1.00**	0.83	**1.00**	**1.00**	**1.00**	0.75
	F1-values	0.86	0.96	**1.00**	0.89	0.96	0.91	0.96	0.96	0.96	0.67

them, the predicted Precision, Recall, and F1 values of MLP for level IV mining area reached 1, the Precision values of KNN, DT, and ET for level II mining area reached 1, and the Precision value of KNN for level III mining area reached 1. Through the above analysis and Fig. 9(a–c), the prediction of mining area of level III and IV is better than that of level I This is related to the fact that there are more noise points in the present data set for grades I and II.

6.2. Discussion

This study conducted an analysis of evaluation indicators for the collected data from goaf areas and compared it with previous research [19]. In comparison to previous studies, the evaluation indicators in this study are more concise, requiring less data collection. This implies a reduction in cost and workload associated with data collection for predicting goaf stability levels.

To reduce the number of indicators while maintaining interpretability, this study employed feature importance and Pearson correlation coefficient for indicator selection and analysis. This approach helps identify the most important and relevant indicators for predicting goaf stability, thereby establishing a more interpretable evaluation indicator system. In contrast, some scholars have employed methods like PCA for dimensionality reduction of indicators, which may result in decreased interpretability of the evaluation indicators.

Furthermore, this study conducted a comparative analysis of multiple machine learning algorithms to determine the most suitable algorithm for predicting goaf stability levels. In contrast to previous studies that solely used a single algorithm [2,21], this research improves scientific rigor and reliability by comparing and evaluating multiple algorithms using various performance indicators. The findings demonstrate that the ET (Extra Trees) algorithm performs best in predicting goaf stability. This discovery is of significant guidance for future predictions of goaf stability classification. By employing machine learning algorithms suitable for this problem, the prediction accuracy of goaf stability can be enhanced, providing a scientific basis for related work.

In conclusion, this study achieved favorable results in predicting goaf stability by streamlining evaluation indicators, employing indicators analysis methods with better interpretability, and comparing multiple machine learning algorithms. These findings are of great importance for further enhancing goaf stability evaluation methods and prediction techniques.

7. Conclusion

1. In this study, a risk level dataset containing 121 groups in the goaf is constructed, and the SMOTE algorithm is used to make the dataset balanced for the problem of unbalanced dataset, RobustScaler is used to standardize the dataset for the case of outliers in the data to enhance the robustness of the dataset.

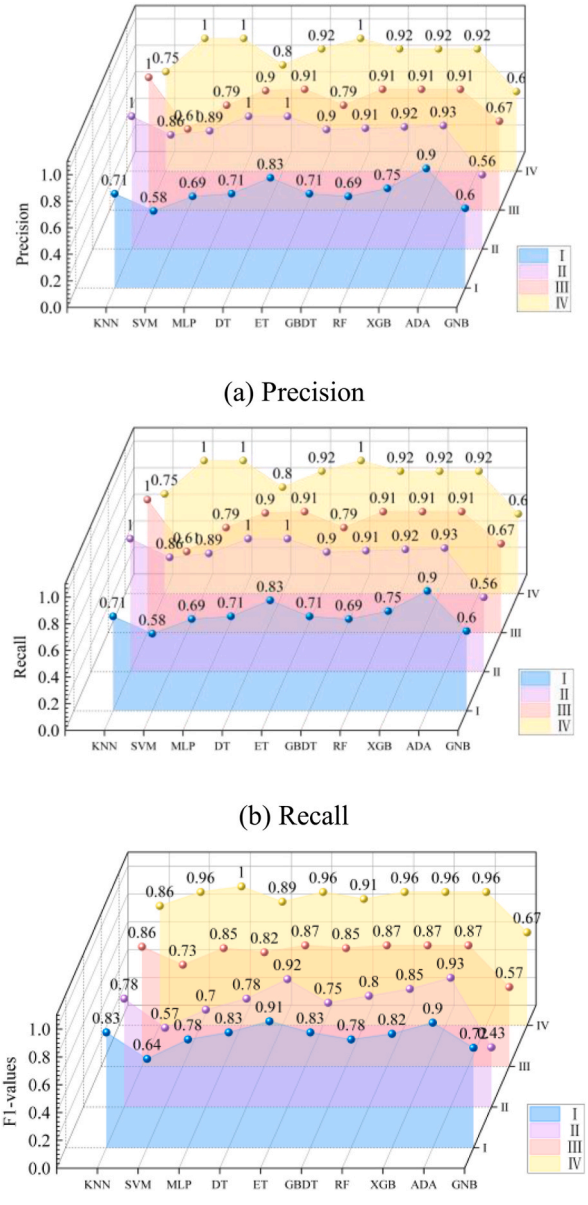


Fig. 9. Accuracy, Recall, and F1 values of different algorithms and different risk levels when Group E features are used (a, b, and c represent the values of Accuracy, Recall, and F1 values, respectively).

2. To solve the problem of redundant evaluation indexes in the goaf, this paper combines and compares various machine learning methods with correlation analysis and feature importance ranking, and comes up with a combination of features applicable to the risk level prediction of the goaf: groundwater, goaf layout, goaf volume, roof exposed area, goaf span height ratio, and mining disturbance.
3. The performance on the test set, the cross-validation results, and the prediction results of each risk level are combined to conclude that the ET, RF and Adaboost algorithms perform well and have better evaluation results for the goaf of risk levels III and IV, with the accuracy of ET reaching 94%, while the performance of SVM, GNB, and other algorithms is not so satisfactory.

This study selected multiple indicators that affect the stability of goaf areas and developed a concise evaluation system for assessing the stability of goaf areas. This evaluation system helps to reduce the workload associated with collecting field data. Ten commonly used machine learning algorithms were applied to predict goaf stability, and through comparative analysis, a more suitable algorithm

for goaf stability prediction was identified. These findings provide a reference for future research on goaf prediction. By utilizing the prediction model proposed in this study, it becomes possible to effectively identify the stability level of goaf areas and take timely measures to ensure the safety of underground mines. The conclusions drawn from this study contribute to the field of underground mining.

However, with the continuous advancement of machine learning techniques and the generation of high-quality data, there is potential for the application of faster and more optimal algorithms as well as larger volumes of high-quality data in this field. These areas represent promising directions for future research.

Author contribution statement

Zhang Bin: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Hu Shaohua: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data.

Li Moxiao: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Fundings

This work has been funded by National Natural Science Foundation of China (grant number is 52209146)

Data availability statement

Data included in article/supp. Material/referenced in article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J. Brodny, M. Tutak, Applying computational fluid dynamics in research on ventilation safety during underground hard coal mining: a systematic literature review, Process Saf. Environ. Protect. 151 (2021) 373–400, https://doi.org/10.1016/j.psep.2021.05.029.

[2] H. Luan, H. Lin, Y. Jiang, Y. Wang, J. Liu, P. Wang, Risks induced by room mining goaf and their assessment: a case study in the Shenfu-Dongsheng mining area, Sustain. Times 10 (2018) 3, https://doi.org/10.3390/su10030631.

[3] H. Li, B. Zhao, G. Guo, J. Zha, J. Bi, The influence of an abandoned goaf on surface subsidence in an adjacent working coal face: a prediction method, Bull. Eng. Geol. Environ. 77 (1) (2018) 305–315, https://doi.org/10.1007/s10064-016-0944-9.

[4] J. Brodny, M. Tutak, A. John, Analysis of influence of types of rocks forming the goaf with caving on the physical parameters of air stream flowing through these gob and adjacent headings, Mechanika 24 (1) (2017) 43–49, https://doi.org/10.5755/j01.mech.24.1.20214.

[5] M. Tutak, J. Brodny, Analysis of influence of goaf sealing from tailgate on the methane concentration at the outlet from the longwall, IOP Conf. Ser. Earth Environ. Sci. 95 (4) (2017), 42025, https://doi.org/10.1088/1755-1315/95/4/042025.

[6] J. Brodny, M. Tutak, Analysis of methane hazard conditions in mine headings, Teh. Vjesn. 25 (1) (2018) 271–276, https://doi.org/10.17559/TV-20160322194812.

[7] Y.X. Hu, X.B. Li, Bayes discriminant analysis method to identify risky of complicated goaf in mines and its application, Trans. Nonferrous Met. Soc. China (English Ed. 22 (2) (2012) 425–431, https://doi.org/10.1016/S1003-6326(11)61194-1.

[8] X. Yan, X. Chen, F. Gong, Instability identification of goaf risky in mines based on distance discriminant analysis method, Adv. Mater. Res. 255–260 (2011) 3740–3743. https://dx.doi.org/10.4028/www.scientific.net/AMR.255-260.3740.

[9] L. Dong, W. Shu, X. Li, Z. Zhou, F. Gong, X. Liu, Quantitative evaluation and case study of risk degree for underground goafs with multiple indexes considering uncertain factors in mines, Geofluids (2017), https://doi.org/10.1155/2017/3271246, 2017.

[10] H. Ma, J. Wang, Y. Wang, Study on mechanics and domino effect of large-scale goaf cave-in, Saf. Sci. 50 (4) (2012) 689–694, https://doi.org/10.1016/j.ssci.2011.08.050.

[11] Y. Sun, X. Zhang, W. Mao, L. Xu, Mechanism and stability evaluation of goaf ground subsidence in the third mining area in Gong Changling District, China, Arabian J. Geosci. 8 (2) (2015) 639–646, https://doi.org/10.1007/s12517-014-1270-9.

[12] X.F. Ao, X.L. Wang, X.B. Zhu, Z.Y. Zhou, X.X. Zhang, Grouting simulation and stability analysis of coal mine goaf considering hydromechanical coupling, J. Comput. Civ. Eng. 31 (3) (2017), https://doi.org/10.1061/(ASCE)CP.1943-5487.0000640.

[13] X. Ao, X. Wang, X. Zhu, Z. Zhou, X. Zhang, Grouting simulation and stability analysis of coal mine goaf considering hydromechanical coupling, J. Comput. Civ. Eng. 31 (3) (2017), 04016069, https://doi.org/10.1061/(asce)cp.1943-5487.0000640.

[14] F. Gao, D. Stead, H. Kang, Y. Wu, Discrete element modelling of deformation and damage of a roadway driven along an unstable goaf - a case study, Int. J. Coal Geol. 127 (2014) 100–110, https://doi.org/10.1016/j.coal.2014.02.010.

[15] Z. Shi, Q. Wang, P. Wang, D. He, Y. Bai, H. You, Time series effect on surface deformation above goaf area with multiple-seam mining, Symmetry 12 (9) (2020), https://doi.org/10.3390/SYM12091428.

[16] J. Zhou, X. Li, H.S. Mitri, Classification of rockburst in underground projects: comparison of ten supervised learning methods, J. Comput. Civ. Eng. 30 (5) (2016), 04016003, https://doi.org/10.1061/(asce)cp.1943-5487.0000553.

[17] C. Qi, A. Fourie, G. Ma, X. Tang, X. Du, Comparative study of hybrid artificial intelligence approaches for predicting hangingwall stability, J. Comput. Civ. Eng. 32 (2) (2018), 04017086, https://doi.org/10.1061/(asce)cp.1943-5487.0000737.

[18] J. Brodny, D. Felka, M. Tutak, "The use of the neuro-fuzzy model to predict the methane hazard during the underground coal mining production process,", J. Clean. Prod. 368 (2022) https://doi.org/10.1016/j.jclepro.2022.133258. July.

[19] P. Dey, S.K. Chaulya, S. Kumar, Hybrid CNN-LSTM and IoT-based coal mine hazards monitoring and prediction system, Process Saf. Environ. Protect. 152 (2) (2021) 249–263, https://doi.org/10.1016/j.psep.2021.06.005.

[20] A.H. Soomro, M.T. Jilani, "Application of IoT and artificial neural networks (ANN) for monitoring of underground coal mines,", in: International Conference on Information Science and Communication Technology, ICISCT), 2020, pp. 1–8, https://doi.org/10.1109/ICISCT49550.2020.9080034, 2020.
[21] Z.S. Wang, K.Z. Deng, "Residual subsidence prediction of abandoned mine goaf based on wavelet support vector machines,", 527, no, in: Natural Resources And Sustainable Development Ii, *Pts 1-4* 524, 1st International Conference on Energy and Environmental Protection (ICEEP 2012), 2012, p. 330. +, https://dx.doi.org/10.4028/www.scientific.net/AMR.524-527.330.
[22] Y. Qin, Z. Luo, J. Wang, S. Ma, C. Feng, Evaluation of goaf stability based on transfer learning theory of artificial intelligence, IEEE Access 7 (2019) 96912–96925, https://doi.org/10.1109/ACCESS.2019.2929533.
[23] Y. Roh, G. Heo, S.E. Whang, A survey on data collection for machine learning: a big data-AI integration perspective, IEEE Trans. Knowl. Data Eng. 33 (4) (2021) 1328–1347, https://doi.org/10.1109/TKDE.2019.2946162.
[24] L. Breiman, Bagging predictors, Mach. Learn. 24 (2) (1996) 123–140, https://doi.org/10.1023/A:1018054314350.
[25] R.E. Schapire, The strength of weak learnability, Mach. Learn. 5 (2) (1990) 197–227, https://doi.org/10.1023/A:1022648800760.
[26] Z. Jin, J. Shang, Q. Zhu, C. Ling, W. Xie, B. Qiang, "RFRSF: employee turnover prediction based on random forests and survival analysis,", Lect. Notes Comput. Sci. 12343 (2020) 503–515, https://doi.org/10.1007/978-3-030-62008-0_35. LNCS.
[27] V.N. Vapnik, The support vector method, Lect. Notes Comput. Sci. 1327 (1997) 264–271, https://doi.org/10.1007/bfb0020166.
[28] N. Friedman, D. Geiger, M. Goldszmidt, Bayesian network classifiers, Mach. Learn. 29 (2–3) (1997) 131–163, https://doi.org/10.1023/a:1007465528199.
[29] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Ann. Stat. 29 (5) (2001) 1189–1232, https://doi.org/10.1214/aos/1013203451.
[30] J.R. Quinlan, Induction of decision trees, Mach. Learn. 1 (1) (1986) 81–106, https://doi.org/10.1007/bf00116251.
[31] G. Biau, Analysis of a random forests model, J. Mach. Learn. Res. 13 (2012) 1063–1095.
[32] J.L. Speiser, M.E. Miller, J. Tooze, E. Ip, A comparison of random forest variable selection methods for classification prediction modeling, Expert Syst. Appl. 134 (Nov. 2019) 93–101, https://doi.org/10.1016/j.eswa.2019.05.028.
[33] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, Mach. Learn. 63 (1) (2006) 3–42, https://doi.org/10.1007/s10994-006-6226-1.
[34] T. Chen, C. Guestrin, "XGBoost: a scalable tree boosting system,", Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. 13–17 (2016) 785–794, https://doi.org/10.1145/2939672.2939785. Augu.
[35] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, J. Comput. Syst. Sci. 55 (1) (1997) 119–139, https://doi.org/10.1006/jcss.1997.1504.
[36] R.E. Schapire, A brief introduction to boosting, IJCAI Int. Jt. Conf. Artif. Intell. 2 (5) (1999) 1401–1406.
[37] Y. Zhang, S. Wang, G. Ji, A comprehensive survey on particle swarm optimization algorithm and its applications, Math. Probl Eng. 2015 (2015), https://doi.org/10.1155/2015/931256.
[38] W.N. Chen, J. Zhang, H.S.H. Chung, W.L. Zhong, W.G. Wu, Y.H. Shi, A novel set-based particle swarm optimization method for discrete optimization problems, IEEE Trans. Evol. Comput. 14 (2) (2010) 278–300, https://doi.org/10.1109/TEVC.2009.2030331.
[39] D. Wang, D. Tan, L. Liu, Particle swarm optimization algorithm: an overview, Soft Comput. 22 (2) (2018) 387–408, https://doi.org/10.1007/s00500-016-2474-6.
[40] Y. Qin, Z. Luo, J. Wang, S. Ma, C. Feng, Evaluation of goaf stability based on transfer learning theory of artificial intelligence, IEEE Access 7 (2019) 96912–96925, https://doi.org/10.1109/ACCESS.2019.2929533.
[41] Wang Chao, Jinping Guo, Liguan Wang, Recognition of goaf risk based on support vector machines method, J. Chongqing Univ. 38 (4) (2015), https://doi.org/10.11835/j.issn.1000-582X.2015.04.013 (in Chinese).
[42] Zijun Li, Wuqing Lin, Chen Yang, Evaluation on risk of goaf based on AGA-BP neural network, Journal of Safety Science and Technology 11 (7) (2015) (in Chinese).
[43] M. Friendly, Comment on 'the generalized pairs plot, J. Comput. Graph Stat. 23 (1) (2014) 290–291, https://doi.org/10.1080/10618600.2013.801777.
[44] R.A.M. Villanueva, Z.J. Chen, "ggplot2: elegant graphics for data analysis, Meas. Interdiscip. Res. Perspect. 17 (3) (2019) 160–167, https://doi.org/10.1080/15366367.2019.1565254, 2nd ed.
[45] X.W. Liang, A.P. Jiang, T. Li, Y.Y. Xue, G.T. Wang, LR-SMOTE — an improved unbalanced data set oversampling based on K-means and SVM, Knowl. Base Syst. 196 (2020), https://doi.org/10.1016/j.knosys.2020.105845.
[46] G. Douzas, F. Bacao, F. Last, "Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE,", Inf. Sci. 465 (1–20) (2018) https://doi.org/10.1016/j.ins.2018.06.056.
[47] D. Edelmann, T.F. Móri, G.J. Székely, On relationships between the Pearson and the distance correlation coefficients, Stat. Probab. Lett. 169 (2021), 108960, https://doi.org/10.1016/j.spl.2020.108960.
[48] T. Phillips, C. GauthierDickey, R. Thurimella, Using transitivity to increase the accuracy of sample-based pearson correlation coefficients, Lect. Notes Comput. Sci. 6263 (2010) 157–171, https://doi.org/10.1007/978-3-642-15105-7_13. LNCS, no. DaWaK.