Mining of Mineral Deposits
Volume 19 (2025), Issue 4, 72-80

MMD JOURNAL / MINING.IN.UA

https://doi.org/10.33271/mining19.04.072

Machine learning model for windage alteration fault diagnosis of mine ventilation system under unbalanced samples

Zhiyuan Shen [1*], Meiling Yan [1], Dan Zhao [2]

[1] *School of Civil Engineering, Shenyang Jianzhu University, Shenyang, China*
[2] *College of Safety Science & Engineering, Liaoning Technical University, Fuxin, China*
**Corresponding author: e-mail szy0207@sjzu.edu.cn*

Abstract

Purpose. The purpose of this paper is to improve the accuracy of resistance variation fault diagnosis in mine ventilation systems under unbalanced datasets.

Methods. Based on WGAN-div, the unbalanced dataset is enhanced to achieve effective expansion of the original samples. A Bagging integrated learning and ResNet deep learning model is integrated to facilitate fault diagnosis of the ventilation system.

Findings. Taking the simple T-shaped ventilation network as an example, fault datasets with unbalance ratios of 1:2, 1:8, 1:10, and 1:20 are constructed. The influence of unbalanced samples on windage alteration fault diagnosis (WAFs) of the ventilation system is deeply analyzed. Taking the ventilation system of Dongshan coal mine as the experimental object, fault diagnosis comparison experiments are conducted using different data augmentation models and classification models. Multiple evaluation indicators, along with t-SNE visualization, are used to assess the validity of the models. The results show that the data generated by the WGAN-div model has a good similarity to the real data. Compared to the GAN, WGAN, and WGAN-GP, the WGAN-DIV is superior. The performance of the ResNet deep learning model has improved significantly.

Originality. This paper conducts research on fault diagnosis of ventilation systems using unbalanced datasets from both the data level and the network system level, effectively addressing the issue of sample imbalance in the actual working conditions of mine ventilation systems.

Practical implications. The proposed method can provide technical support for the application of intelligent ventilation, enhancing both the reliability of monitoring and the overall safety performance of mine ventilation systems.

Keywords: *mine ventilation system, fault diagnosis, unbalanced samples, generative adversarial network, Bagging-ResNet*

1. Introduction

The stable operation of the mine ventilation system is a crucial factor in ensuring the safety of mine production, the orderly operation of equipment, and disaster prevention and reduction. With the advancement of underground mining work, the expansion of roadway branches is increasing, and the difficulty of ventilation system management is also increasing. However, air leakage, tunnel caving, damper opening, damage to ventilation facilities, and other faults in the ventilation system can cause changes in wind resistance along the roadway. The fault that causes the permanent change of the roadway wind resistance is called the WAFs [1]. The WAFs will cause a change in underground air flow, leading to the accumulation of dust and gas, and reducing the stability of the mine ventilation system and its ability to resist disasters. The topological relationship of the underground ventilation network is complicated. When the wind resistance of a branch changes, the air volume of itself and other branches will also change accordingly; therefore, the data monitored by the wind speed sensor can only indicate the change of the air volume of the roadway where the sensor is located, but it cannot be determined which roadway is faulty. Therefore, determining the fault location in a timely and accurate manner has become a challenging problem to solve in coal mines [2]-[4]. Applying a machine learning algorithm to realize intelligent fault diagnosis of the ventilation system and help intelligent management of mine ventilation is the key to this research.

With the development of big data, industrial Internet, artificial intelligence and other technologies, fault diagnosis technology has matured in different engineering fields such as power system [5], [6], aerospace [7], [8], mechanical equipment [9], wind power generation [10], [11], automobile [12], [13], heating system [14], water supply and distribution system [15] and so on.

In 2018, Liu et al. [16], [17] utilized air volume as an input feature. They employed the Support Vector Machine algorithm to identify the fault location and quantity of the mine ventilation system, thereby pioneering the application of machine learning to fault diagnosis in mine ventilation systems. In 2020, an unsupervised fault diagnosis model for mine ventilation systems was developed using a genetic algorithm, which eliminates the need for training samples and significantly enhances diagnostic performance. Huang et

Received: 4 July 2025. Accepted: 24 November 2025. Available online: 30 December 2025

Mining of Mineral Deposits. ISSN 2415-3443 (Online) | ISSN 2415-3435 (Print)

al. employed a Kalman filter model to preprocess the wind speed monitoring data of the mine. They proposed an unsupervised learning fault diagnosis model for the mine ventilation system, based on a hybrid coding algorithm, to simultaneously diagnose fault location and fault volume [1], [18]. Zhou et al. optimized the parameters of the SVM model for fault diagnosis of the mine ventilation system using an improved genetic algorithm, effectively avoiding the model overfitting problem [19]. Ni et al. proposed a ventilation system fault diagnosis method based on Random Forest and Decision Tree. They confirmed that the Random Forest model is superior to the Decision Tree model [20], [21]. Zhang et al. selected the SVM algorithm, the ANN algorithm, and the RF algorithm to conduct a comparative analysis of the fault diagnosis of the mine ventilation system, and the results showed that the ANN algorithm had higher accuracy [22]. Zhao et al. used the Daming coal mine as the subject of their research. They applied the improved SVM algorithm to the fault diagnosis of the ventilation system in the fault roadway range database, thereby reducing the fault location range and improving the sample training efficiency [23]. In 2022, Wang et al. studied the identification algorithm when multiple branches of the mine ventilation system failed at the same time, and built a machine learning model based on multi-label K-nearest neighbor, which was the first proposed method to solve the rapid diagnosis when multiple locations of the mine ventilation system failed [24]. Liu et al. applied four machine learning algorithms to fully evaluate the performance of the fault diagnosis model for the mine ventilation system. They determined the superiority of the KNN algorithm and the DT algorithm. Meanwhile, the influence of four factors, sample dispersion, sample number, input feature and feature number on the generalization performance of the fault diagnosis model is analyzed, which provides a reference for the establishment of a machine learning model for WAFs of the ventilation system [25], [26].

Currently, a fault diagnosis model for the mine ventilation system is being established based on relatively complete data sets. However, in the actual ventilation system failure situation, a complete data set cannot be obtained. How to carry out fault diagnosis of the ventilation system in the case of unbalanced samples is a serious challenge. In view of this, the authors conduct research on the fault diagnosis of WAFs using unbalanced samples from both the data level and the network system level. In this research, a Wasserstein distance for GANs (WGAN-div) model is constructed to enhance the original data and reconstruct a balanced dataset. The WAFs of the ventilation system are realized by integrating the Bagging ensemble learning model and the ResNet deep learning model. This research provides technical support for the practical application of intelligent diagnosis technology in the mining industry.

2. Methods

2.1. Unbalanced analysis of the ventilation system fault sample

In actual working conditions of mine ventilation systems, the roadway containing ventilation structures, the mining face, the main windway, the intersection of ventilation branches, and other positions are prone to failure. These roadways generate more fault data, while other roadways generate less fault data. There is a significant gap in the number of fault samples generated by each branch, resulting in a data imbalance problem. The unbalanced data set of fault branches in the mine ventilation system can be described by Equation (1):

$$\begin{cases} S_{m+n} = \{X_m, Y_n\} \\ X_m = \{x_i \mid i = 1,2,\cdots,m\}, \\ Y_n = \{y_i \mid i = 1,2,\cdots,n\} \end{cases} \tag{1}$$

where:

X_m – the minority class fault branch data set;

Y_n – the majority class fault branch data set;

S_{m+n} – the unbalanced data set of fault branches of the ventilation system;

x_i and y_i – the i^{th} sample data in each data set;

m – the number of minority class samples;

n – the number of majority class samples.

2.2. The WGAN-div model

The Generative Adversarial Network model can generate new sample data, thereby adjusting the balance between X_m and Y_n. The GAN model is primarily composed of two parts: the discriminator and the generator. However, the traditional GAN model is prone to instability during training [27]. In 2017, Arjovsky et al. developed a Wasserstein divergence for GANs (WGAN) model to address the issue of gradient disappearance during the training of traditional GAN models [28]. However, during WGAN training, it is typically necessary to keep the absolute value of the gradient below a fixed threshold. Literature [29] proposes a WGAN-GP model with penalty factors to ensure Lipschitz continuity between generated samples and real samples, but there is no theoretical basis for this scheme. Literature [30] proposes a WGAN-div model that does not require Lipschitz constraints and proves its superiority both theoretically and in application. Based on previous studies, this paper selects the WGA-div data enhancement model, and the loss functions are shown in Equations (2) and (3):

$$L_G = -E_{G(z)\sim P_G}\left[D\left(G(z)\right)\right]; \tag{2}$$

$$L_D = E_{G(z)\sim P_G}\left[D\left(G(z)\right)\right] - E_{x\sim M_r} \times \\ \times \left[D(x) - kE_{\hat{x}\sim p_u}\left\|\nabla_{\hat{x}} D(\hat{x})\right\|^p\right], \tag{3}$$

where:

L_G – the generator loss function;

L_D – the discriminator loss function;

$E_{G(z)\sim PG}$ – the expected function of generator noise;

$E_{\hat{x}\sim p_u}$ – the expectation function of interpolating $\hat{x}$;

$\hat{x}$ – the random interpolation between the generated sample and the real sample;

P_u – is the distribution of interpolation $\hat{x}$;

k and *powers* of the norm (according to previous studies and experimental tests, k is 2 and p is 6 in this paper).

To prevent the problem of gradient disappearance or network degradation during WGAN-div model training, identity mapping residuals are added to both the discriminator and generator. In this paper, the WGAN-div model with residual blocks is applied to enhance the unbalanced samples of ventilation system monitoring data. The number of minority samples in the ventilation system sample fault data set is adjusted from m to. The balanced data set $S' = \{X'_r, Y_n\}$ is further obtained, which is the balanced minority sample data set.

2.3. Bagging-ResNet

The wind speed monitoring data of the mine ventilation system has a large dimension, which belongs to the high-dimensional unbalanced data. The ResNet is well-suited for processing classification problems involving high-dimensional data. Still, its performance is significantly influenced by the number of neurons, connection mode, number of network layers, and initial weights. This paper attempts to combine Bootstrap aggregating (Bagging) and ResNet to build the Bagging-ResNet model, utilizing an integrated learning model with strong generalization ability to mitigate the instability of a single residual network. Thus, the accuracy of WAFs diagnosis of the mine ventilation system is improved. The wind speed data is taken as the input of the Bagging-ResNet classification model, and the fault branch number is taken as the output of the classification model. The specific process is as follows: the sample data set is sampled using Bootstrap, and K sample subsets are obtained. ResNet models are then built. The test data are input into the K models to obtain K results. The weighted average formula is used to obtain the final classification results as shown in Equation (4):

$$y_{end} = \frac{1}{K}\sum_{i=1}^{K} y_i , \tag{4}$$

where:

y_{end} – the final classification of the test sets;

y_i – the result of classification of the test set by the i^{th} ResNet model;

y_{end} and y_i – both probability values.

2.4. Overall structure and flow of the WAFs diagnosis model

The overall framework of WAFs diagnosis based on WGAN-div-Bagging-ResNet is shown in Figure 1. The specific process is as follows. The intelligent mine ventilation simulation system (IMVS) is used to simulate ventilation system faults, and an unbalanced data set O is constructed, which is divided into a test set O_{st} and a train set O_{in}.

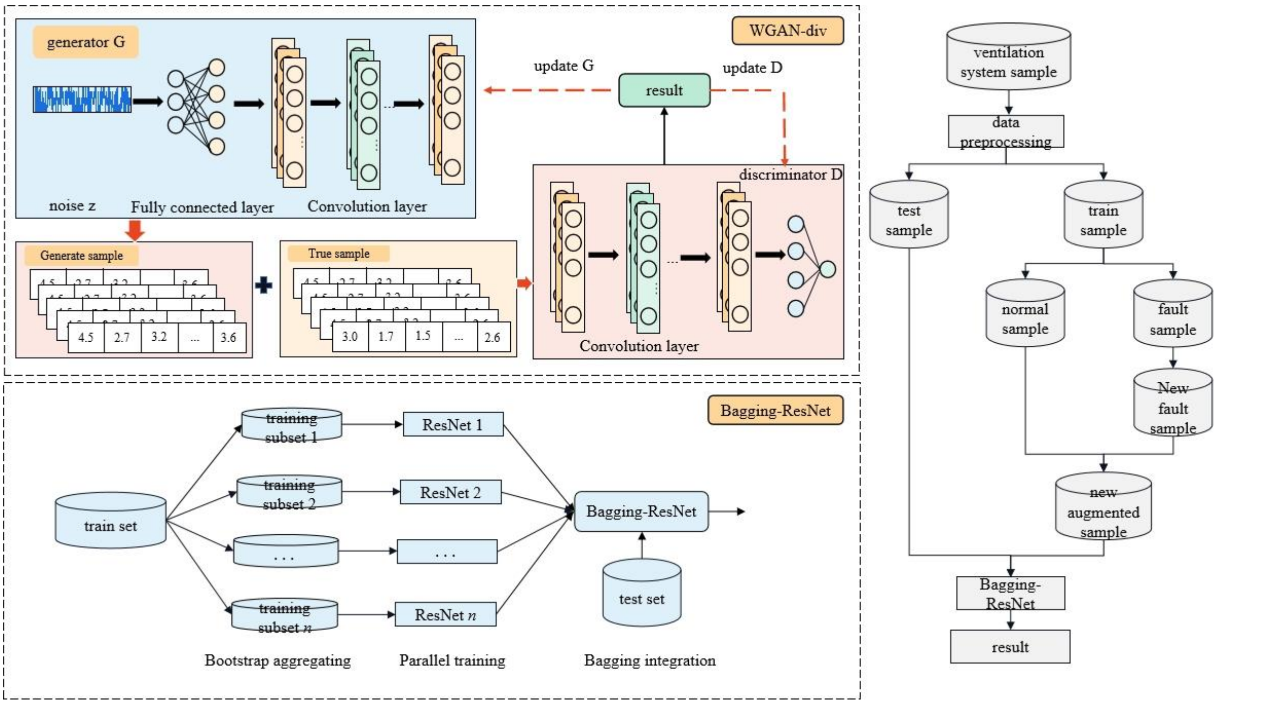

Figure 1. WGAN-div-Bagging-ResNet model architecture

The WGAN-div model is applied to the train set O_{in} for data enhancement processing, and a new fault sample O_n is generated, which is added to the original train set O_{in} to synthesize a new augmented sample O_{ex}. The Bagging-ResNet model is trained by the balanced augmented sample set O_{ex}, and the trained WAFs diagnosis model is obtained. The test set O_{st} is input into the trained Bagging-ResNet model for diagnosing ventilation system faults.

2.5. Evaluation index

The evaluation of the WAFs diagnosis of the ventilation system multi-classification model is usually based on a binary classification confusion matrix. For a multi-classification problem with unbalanced samples, it isn't easy to achieve an accurate evaluation of classification results. Therefore, *Re*, *Pr*, *G-mean*, and *F*1 are added to evaluate the WAFs diagnosis model comprehensively. The definitions of each indicator are shown in Equations (5)-(8):

$$Pr = \frac{\sum_{i=1}^{N} T_{pi}}{\sum_{i=1}^{N} T_{pi} + \sum_{i=1}^{N} F_{pi}} ; \tag{5}$$

$$Re = \frac{\sum_{i=1}^{N} T_{pi}}{\sum_{i=1}^{N} T_{pi} + \sum_{i=1}^{N} F_{Ni}} ; \tag{6}$$

$$F1 = 2 \cdot \frac{Pr \cdot Re}{Pr + Re} ; \tag{7}$$

$$G\text{-}mean = \sqrt{\frac{\sum_{i=1}^{N} T_{pi}}{\sum_{i=1}^{N} T_{pi} + \sum_{i=1}^{N} F_{Ni}} \cdot \frac{\sum_{i=1}^{N} T_{Ni}}{\sum_{i=1}^{N} T_{Ni} + \sum_{i=1}^{N} F_{pi}}} , \tag{8}$$

where:

N – the number of branches of the ventilation network in the input model;

T_{Pi} – the true positive of the i^{th} faulty branch;

T_{Ni} – the true negative of the i^{th} faulty branch;

F_{Pi} – the false positive of the i^{th} faulty branch;

F_{Ni} – the false negative of the i^{th} faulty branch.

3. Results and discussion

3.1. Influence of unbalanced data on WAFs diagnosis

Taking the simple T-type ventilation network shown in Figure 2 as an example, WAF diagnosis experiments are designed under different unbalance ratios. The ventilation network has 10 branches and 8 nodes. Branches e_1 and e_{10} are the inlet and return air branches, respectively. There is an air window adjustment facility located in the e_5. The fan characteristic equation is $H(q) = 1035.9 + 51.8q - 0.43q^2$.

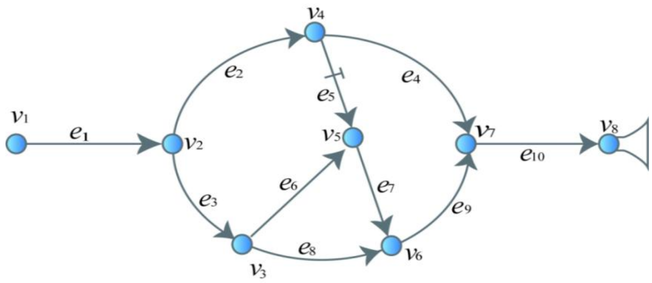

Figure 2. T ventilation network diagram

The IMVS is used to simulate branch faults, and six sets of unbalanced data are generated according to different imbalance ratios. Six sets of experimental schemes are then constructed. In the T-type network, e_5 is equipped with an air window, which is more prone to failure than other branches; therefore, the unbalance ratio can be adjusted by increasing the number of failures of e_5. H is set to 1:2, 1:8, 1:10, and 1:20, respectively. The number of simulated failures of the e_5 is set to 100, 250, 400, 500, and 1000. Additionally, the number of fault samples for a few classes is set to 50. In this experiment, air volume and wind pressure are used as the input characteristics of the model. The corresponding air volume characteristic experiment schemes are denoted T_1, T_2, T_3, and T_4, respectively, and the wind pressure characteristic experiment schemes are denoted D_1, D_2, D_3, and D_4, respectively. Different classification models are used to identify the fault branches of the ventilation system. The classification models are selected as SVM, KNN, DT, and MLP.

3.1.1. The result of WAFs diagnosis of the wind volume characteristic

The optimal hyperparameters for each experimental model under varying wind volume characteristics are determined using the 50% cross-validation method, as shown in Table 1. The definitions of each parameter are provided in Table 2.

Figure 3 shows that when wind volume is used as an input feature and SVM is used as a classification model, the average index values of Pr, Re, $F1$, and G-$mean$ of T_1 are 0.9, 0.86, 0.88, and 0.84, respectively. The average index values of Pr, Re, $F1$, and G-$mean$ of T_2 are 0.85, 0.71, 0.77, and 0.69, respectively. The average index values of Pr, Re, $F1$, and G-$mean$ of T_3 are 0.77, 0.63, 0.68, and 0.62, respectively. The average index values of Pr, Re, $F1$, and G-$mean$ of T_4 are 0.68, 0.49, 0.55, and 0.59, respectively. With the increase in data imbalance, the evaluation index values of the SVM classification model showed a gradual decline.

Table 1. WAFs results of each model with the feature of wind volume

Algorithm	Parameter	T_1	T_2	T_3	T_4
SVM	g	0.241	0.121	0.012	0.321
	c	0.1	0.1	0.2	0.2
	K	RBF	RBF	RBF	RBF
KNN	kn	1	1	2	2
	lp	6	4	1	10
DT	M_f	0.2	0.6	0.6	0.4
	N_{leaf}	2	5	5	3

Table 2. Definition of each classification model parameter

Parameter	Definition
c	Penalty coefficient
g	Nuclear parameter
K	Kernel function
N_{leaf}	The minimum number of samples required for leaf nodes
kn	The number of nearest neighbors
lp	Minkowski distance power parameters
M_f	The best segmentation ratio of features

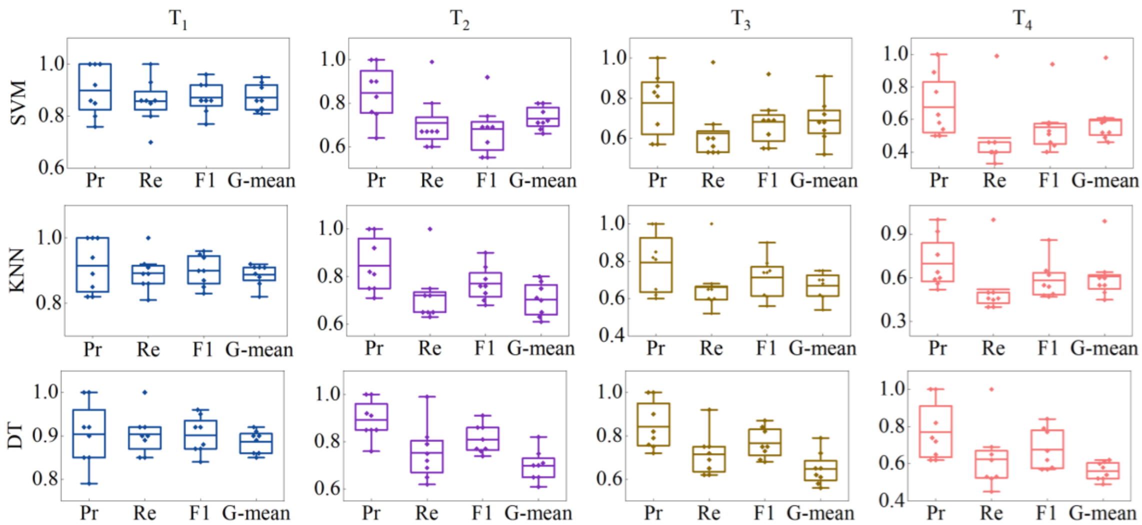

Figure 3. Evaluation indices of each model with the feature of wind volume under the imbalanced data

Compared with experiment T_1 (data unbalance ratio 2:1) and experiment T_4 (data unbalance ratio 20:1), the *Pr*, *Re*, *F*1, and *G-mean* values decreased by 22, 38, 33, and 25%, respectively. Evaluation indices of KNN and DT classification models show the same trend as SVM, *Pr*, *Re*, *F*1, and *G-mean* values of experiment T_4 in the KNN model decreased by 22, 37, 32, and 29%, respectively, compared with experiment T_1. In the DT model, the *Pr*, *Re*, *F*1, and *G-mean* values of experiment T_4 decreased by 13, 28, 23, and 27%, respectively, compared with experiment T_1.

It can be seen that unbalanced data significantly affects the overall performance of the model, resulting in a reduced index score. Furthermore, unbalanced data leads to more missed judgments and incorrect predictions in each model.

3.1.2. The result of WAFs diagnosis of the wind pressure characteristic

The best hyperparameters of each experimental model under wind pressure characteristics are shown in Table 3, and the definitions of each parameter are shown in Table 2. WAFs diagnosis evaluation indexes of three machine learning algorithms are shown in Figure 4.

Table 3. WAFs results of each model with the feature of wind pressure

Algorithm	Parameter	T_1	T_2	T_3	T_4
SVM	*g*	0.12	0.154	0.35	0.51
	c	0.1	0.1	0.02	0.1
	K	RBF	RBF	RBF	RBF
KNN	*kn*	2	1	2	3
	lp	6	4	5	10
DT	M_f	0.2	0.3	1.2	0.4
	N_{leaf}	2	5	3	3

As shown in Figure 4, with wind pressure as the input feature, the *Pr*, *Re*, *F*1, and *G-mean* values of T_4 decreased by 27, 30, 28, and 30%, respectively, compared to experiment T_1, using the SVM algorithm. The *Pr*, *Re*, *F*1, and *G-mean* values of T_4 decreased by 29, 25, 27, and 29%, respectively, using the KNN algorithm. The *Pr*, *Re*, *F*1, and *G-mean* values of T_4 decreased by 25, 23, 28, and 27%, respectively, using the DT algorithm. This indicates that the WAFs diagnosis results, which use wind pressure as the input feature, are also affected by the degree of sample imbalance.

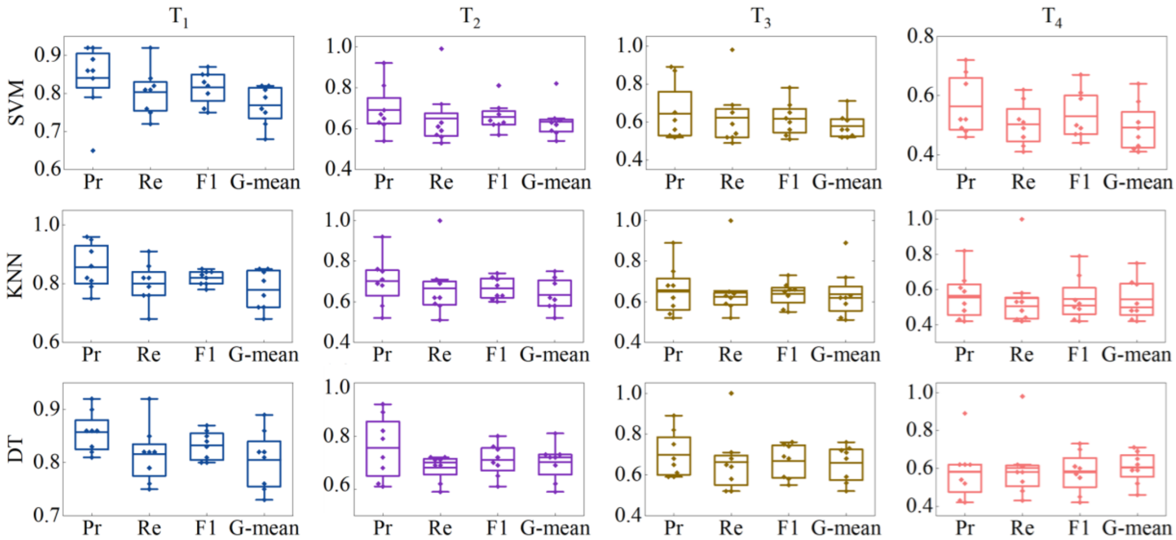

Figure 4. Evaluation indices of each model with the feature of wind pressure under the imbalanced data

Compared with the indices of each model under the cha-racteristics of wind pressure, it can be seen that the indices under the characteristics of wind volume are lower, indica-ting that the characteristics of wind volume are more suitable for the WAFs diagnosis of the mine ventilation system.

To sum up, the conventional machine learning classifier builds a model based on the rules derived from a large amount of data, while ignoring the characteristics of data from other branches. As a result, it is easy to misdiagnose the faults of other branches as the branches of the majority class during classification. As the imbalance ratio increases, the proportion of fault samples that are misjudged gradually increases. If the fault branch is misidentified as another branch, the optimal maintenance time for the mine ventilation tunnel will be missed. This further illustrates the harm of unbalanced samples to the WAFs diagnosis model. The research is both necessary and practical.

3.2. Large mine test

3.2.1. WAFs data preparation

Taking the ventilation system of Dongshan coal mine as an example, a diagnosis test is conducted using the unbalanced sample WAFs. The ventilation mode of the mine is diagonal, and the ventilation network of the mine is shown in Figure 5. The number of branches is 96, the number of nodes is 84, and the total inlet air volume is 14394 m^3/min. The branches corresponding to the four inlet shafts are e_2, e_1, e_{23}, and e_5, respectively. The branch numbers for installing the damper are e_{47}, e_{85}, e_{28}, e_{86}, e_{48}, e_{78}, e_{22}, e_7, e_{30}, e_{38}, e_{29}, e_{19}, e_{65}, e_{52}, e_{84}, and e_{33}. The branch numbers for installing wind windows are e_{10}, e_{83}, e_{24}, e_{13}, and e_{93}. IMVS is used to simulate branch faults. These branches of installed structures are simulated 200 times with faults. The other branches are simulated with 10 times the faults. A total of 5120 fault samples are obtained. The data imbalance ratio is 20:1. According to the experimental results in Section 4.1, wind speed is selected as the input feature of the model in this paper, and some data are shown in Table 4.

Table 4. Fault sample set in the production mine

Samples	v'_1	v'_2	v'_3	v'_4	v'_5	v'_6	v'_7	v'_8	v'_9	v'_{10}	v'_{11}	v'_{12}	v'_{13}	v'_{14}	v'_{15}	e'_i
1	3.6	5.6	7.6	4.2	3.9	4.1	2.6	10.4	4.7	4.2	3.2	3.8	9.6	4.2	3.2	e_{85}
2	3.6	5.4	7.8	4.1	3.9	4.2	2.4	10.2	4.8	4.2	3.1	3.9	9.5	4.1	3.3	e_{85}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
201	3.5	3.6	7.9	4.1	3.8	7.1	2.6	3.6	4.7	4.3	3.2	8.0	4.3	4.2	3.1	e_{33}
202	3.4	3.5	7.8	4.2	3.7	7.2	2.5	3.7	4.6	4.2	3.1	7.9	4.4	4.2	3.2	e_{33}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
5119	6.5	5.6	7.6	4.4	3.6	4.4	2.6	0.7	4.5	4.3	3.2	3.8	0.5	4.0	3.1	e_{92}
5120	6.4	5.6	7.8	4.3	3.6	4.3	2.5	0.6	4.6	4.4	3.3	3.9	0.3	4.2	3.0	e_{92}

Where v'_i is the wind speed of each branch (m/s), e'_i is the faulty branch. The fault sample data, after standardized processing, is divided into training samples and test samples in a ratio of 7:3.

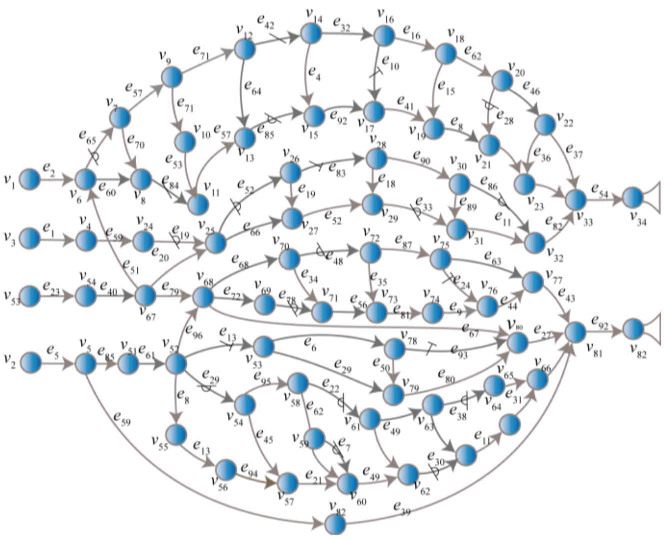

Figure 5. Ventilation network of Dongshan coal mine

3.2.2. Validation of WGAN-div

To verify the effectiveness of WGAN-div in processing unbalanced data from the ventilation system, the original fault samples are processed by the following models: original data set O_{in}, GAN, WGAN, WGAN-gp, and WGAN-div. This allows for the creation of a combined sample set, O_{ex}, to achieve data balance. The Bagging-ResNet algorithms are then chosen for classification. The test results are presented in Table 5, which shows the mean and standard deviation of the results from ten runs (the optimal results are highlighted in bold).

It can be seen from Table 5 that, compared with the original data set, after the enhancement of WGAN-div, ACC increased by 16.1%, *Re* increased by 15.8%, *Pr* increased by 16%, *G-mean* increased by 15.8%, and *F*1 increased by 16.2%, respectively. It demonstrates that utilizing the WGAN-div model to enhance the unbalanced fault data can effectively improve the quality of the original data and enhance the discriminant performance of the classifier. After using GAN, WGAN, and WGAN-GP for data enhancement, although the accuracy and G-mean indexes are increased, the classification model's ability is enhanced; however, there is no noticeable improvement in *F*1.

Table 5. Experimental results of different data enhancement methods

Method	ACC	*Re*	*Pr*	*G-mean*	*F*1
O_{in}	0.810 ± 0.012	0.814 ± 0.06	0.811 ± 0.102	0.814 ± 0.071	0.808 ± 0.12
GAN	0.912 ± 0.025	0.910 ± 0.014	0.911 ± 0.047	0.916 ± 0.025	0.814 ± 0.015
WGAN	0.920 ± 0.021	0.943 ± 0.017	0.724 ± 0.02	0.934 ± 0.024	0.813 ± 0.01
WGAN-gp	0.931 ± 0.018	0.936 ± 0.017	0.845 ± 0.054	0.947 ± 0.02	0.892 ± 0.065
WGAN-div	0.971 ± 0.006	0.972 ± 0.121	0.971 ± 0.014	0.972 ± 0.039	0.97 ± 0.041

The analysis reveals that the model expands new fault samples of poor quality, which impacts the classification model's ability to discriminate between fault branch diagnoses. Compared with GAN, WGAN, and WGAN-gp, the WGAN-div model has the highest evaluation indexes, and the scores of ACC, *Re*, *Pr*, *G-mean*, and *F*1 are 97.1, 97.2, 97.1, 97.1 and 97%, respectively, which significantly improves the classification model's ability to identify fault branches. The superiority of the proposed WGAN-div in processing unbalanced data is verified.

The visualization analysis of the sample generation from the WGAN-div model is performed by applying the t-SNE (t-Stochastic Neighbor Embedding) algorithm. Figure 6 shows the distribution between the generated samples and the real samples of the model for iteration *N* values of 0, 100, 200, 500, 800, and 1000, respectively. Figure 7 shows the change in the model loss function.

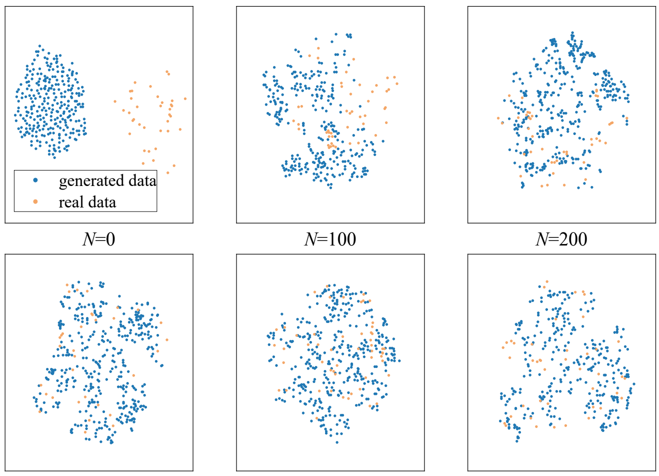

Figure 6. t-SNE dimension reduction data visualization

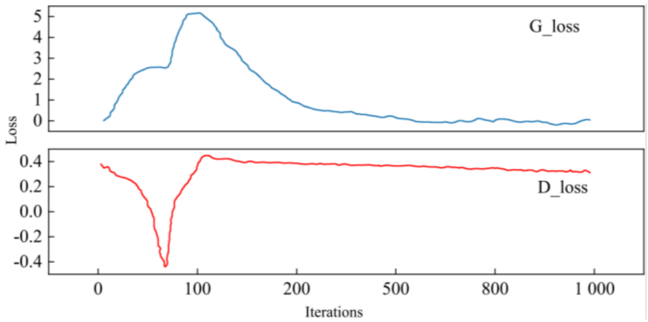

Figure 7. WGAN-div loss function

According to Figures 6 and 7, as the number of iterations increases, the loss function of the WAGN-div model converges steadily and becomes stable gradually. The distribution of the generated new sample data and the real data gradually blends, and the generated data exhibits a high similarity to the real data, with the quality of the generated data continually improving.

3.2.3. Validation of Bagging-ResNet

To demonstrate the superiority of Bagging-ResNet, the following classical ensemble learning classification models are selected for comparison: CBT, LGB, and GBDT. WGAN-div processes the original sample. In addition, the RF of the ventilation system WAFs diagnosis proposed in reference [31] is also included in the comparative test of this paper. To more intuitively reflect the advantages and disadvantages of the classifier, this paper introduces the receiver operating characteristic curve (ROC) and the area under the ROC curve (AUC) for evaluation. The closer the ROC curve is to the vertical axis, the larger the AUC value, and the better the classifier's performance is. If the ROC curve lies below the $y = x$ line, the classification effect is highly unsatisfactory. The evaluation indicators of each model are shown in Figure 8. The ROC curve and AUC values are shown in Figure 9.

It can be seen from Figures 8 and 9 that all evaluation indices of the Bagging-ResNet model are the highest, and the ROC curve is closer to the vertical axis. From the perspective of comprehensive evaluation indicators, compared with RF, CBT, LGB, and GBDT, *Re* increased by 1, 4.5, 3.2, and 4%, respectively, *Pr* increased by 0.8, 4, 3.9, and 5.1%, respectively, *G-mean* increased by 1.1, 5.2, 2.7, and 4.1%, *F*1 increased by 1, 4.4, 3.7, and 4.4%, AUC increased by 1.4, 6.4, 4.1, and 3.5%, respectively.

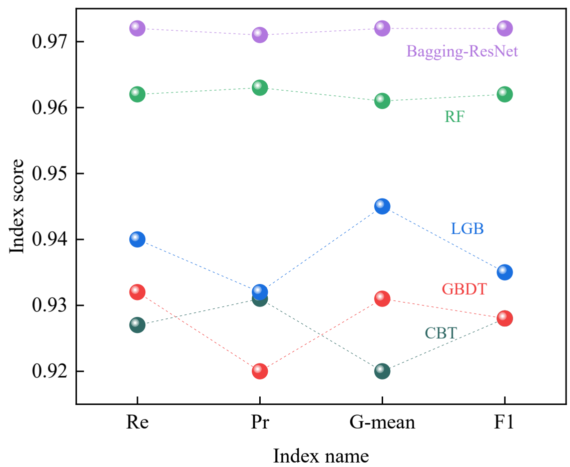

Figure 8. Evaluation index values of each model

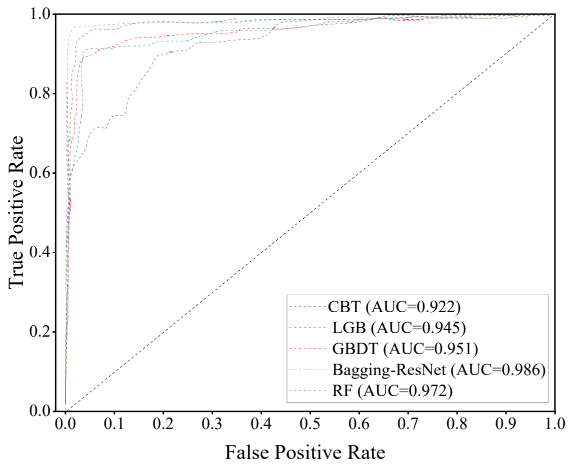

Figure 9. ROC curve and AUC value of each model

The weak classifiers integrated with RF, CBT, LGB, and GBDT are shallow networks, and their learning ability is significantly inferior to that of Bagging-ResNet. In general, the Bagging-ResNet model is suitable for the diagnosis of WAFs in mine ventilation systems. The integration of ResNet with Bagging yields stronger generalization of the model and better performance than other models in terms of *Re*, *Pr*, *G-mean*, *F*1, and AUC.

3.3. Summary of key findings and future research directions

The experimental results demonstrate that the proposed WGAN-div-Bagging-ResNet framework significantly enhances the diagnostic performance for windage alteration faults (WAFs) under imbalanced sample conditions. Compared to conventional machine learning models such as SVM, KNN, and DT, the integrated approach exhibits superior robustness and accuracy, particularly as the imbalance ratio increases. This improvement can be attributed to the synergistic combination of high-quality data augmentation and a deep ensemble classifier, which collectively mitigate the bias toward majority classes and enhance feature learning from limited fault samples. The superior performance of WGAN-div over other generative models (GAN, WGAN, WGAN-gp) lies in its theoretical and structural advantages. The data generated by WGAN-div blends seamlessly with real samples, indicating high distributional similarity. This ensures that the augmented dataset preserves the underlying physical characteristics of ventilation system faults, thereby improving the discriminative capability of the classifier.

The proposed method offers a viable solution for real-world mine ventilation systems, where fault data are often scarce and unevenly distributed. By effectively balancing the dataset and enhancing model generalizability, the framework reduces misdiagnosis and missed alarms. This approach aligns with the trend toward intelligent ventilation management, providing a scalable tool for operational safety in complex mining environments.

Currently, the fault diagnosis of mine ventilation systems primarily relies on simulation methods to generate simulated data for training intelligent diagnostic models. A critical direction for future research is, therefore, the development of precise and rapid testing technologies for key ventilation parameters. Success in this area would enable the collection of high-fidelity, real-condition datasets. Training intelligent diagnostic models on such data would significantly enhance their accuracy, reliability, and generalization capability for direct field deployment.

4. Conclusion

A simple T-ventilation network is taken as an example to illustrate the impact of an unbalanced sample on the WAFs diagnosis model. The unbalanced data of mine ventilation system will degrade the classification performance of traditional machine learning models and even cause failures. The wind volume characteristic is more suitable for the diagnosis of mine ventilation systems using WAFs.

Fault diagnosis tests and t-SNE visualization results indicate that the WGAN-div model with residual blocks can generate high-quality new data, thereby expanding the sample set. The scores of ACC, *Re*, *Pr*, *G-mean*, and *F*1 of WGAN-div are 97.1, 97.2, 97.1, 97.1 and 97%, respectively. Compared to other data enhancement models, WGAN-div has more advantages in handling unbalanced samples.

With the help of the Bagging integration idea, the Bagging-ResNET model is built. By integrating multiple ResNet classifiers, various learning models, including RF, CBT, LGB, and GBDT, are introduced for comparison with ResNet.The comprehensive evaluation index scores of Bagging-ResNet are more favorable than those of other integrated models.

Author contributions

Conceptualization: ZS; Data curation: MY; Formal analysis: ZS; Resources: DZ; Software: MY; Visualization: DZ; Writing – original draft: MY; Writing – review & editing: ZS. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the Basic Scientific Research Project of the Department of Education of Liaoning Province (Grant No. LJ212510153045) and the National Natural Science Foundation of China (Grant No. 52374202).

Acknowledgements

Special thanks are extended to the technical teams of Shenyang Jianzhu University and Liaoning Technical University for their valuable assistance in data collection and simulation work.

Conflicts of interest

The authors declare no conflict of interest.

Data availability statement

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

References

[1] Huang, D., Liu, J., & Deng, L. (2020). A hybrid-encoding adaptive evolutionary strategy algorithm for windage alteration fault diagnosis. *Process Safety and Environmental Protection*, *136*, 242-252. https://doi.org/10.1016/j.psep.2020.01.037

[2] Chen, J., Zhang, L., Li, Y., Shi, Y., Gao, X., & Hu, Y. (2022). A review of computing-based automated fault detection and diagnosis of heating, ventilation and air conditioning systems. *Renewable and Sustainable Energy Reviews*, *161*, 112395. https://doi.org/10.1016/j.rser.2022.112395

[3] Fan, D., Li, C., & Yan, Z. (2020). Overall architecture and core scenario of a smart coal mine incorporating 5G technology ecology. *Journal of China Coal Society*, *45*(06), 1949-1958.

[4] Jia, J., Jia, P., & Li, Z. (2020). Theoretical study on stability of mine ventilation network based on sensitivity analysis. *Energy Science & Engineering*, *8*(8), 2823-2830. https://doi.org/10.1002/ese3.699

[5] Elsisi, M., Tran, M., Mahmoud, K., Mansour, D.E.A., Lehtonen, M., & Darwish, M.M.F. (2022). Effective IoT-based deep learning platform for online fault diagnosis of power transformers against cyberattacks and data uncertainties. *Measurement*, *190*, 110686. https://doi.org/10.1016/j.measurement.2021.110686

[6] Vaish, R., Dwivedi, U.D., Tewari, S., & Tripathi, S.M. (2021). Machine learning applications in power system fault diagnosis: Research advancements and perspectives. *Engineering Applications of Artificial Intelligence*, *106*, 104504. https://doi.org/10.1016/j.engappai.2021.104504

[7] Zhu, M., Liu, Q., & Liu, X. (2020). Fault detection method for avionics based on LMKL and OC-ELM. *Systems Engineering and Electronics*, *42*(06), 1424-1432.

[8] Zhao, Y.-P., & Chen, Y.-B. (2022). Extreme learning machine based transfer learning for aero engine fault diagnosis. *Aerospace Science and Technology*, *121*, 107311. https://doi.org/10.1016/j.ast.2021.107311

[9] He, M., & He, D. (2017). Deep learning based approach for bearing fault diagnosis. *IEEE Transactions on Industry Applications*, *53*(3), 3057-3065. https://doi.org/10.1109/tia.2017.2661250

[10] Pérez-Pérez, E.J., López-Estrada, F.R., Puig, V., Valencia-Palomo, G., & Santos-Ruiz, I. (2022). Fault diagnosis in wind turbines based on ANFIS and Takagi-Sugeno interval observers. *Expert Systems with Applications*, *206*, 117698. https://doi.org/10.1016/j.eswa.2022.117698

[11] Habibi, H., Howard, I., & Simani, S. (2019). Reliability improvement of wind turbine power generation using model-based fault detection and fault tolerant control: A review. *Renewable Energy*, *135*, 877-896. https://doi.org/10.1016/j.renene.2018.12.066

[12] Chakrapani, G., & Sugumaran, V. (2023). Transfer learning based fault diagnosis of automobile dry clutch system. *Engineering Applications of Artificial Intelligence*, *117*, 105522. https://doi.org/10.1016/j.engappai.2022.105522

[13] Jegadeeshwaran, R., & Sugumaran, V. (2015). Fault diagnosis of automobile hydraulic brake system using statistical features and support vector machines. *Mechanical Systems and Signal Processing*, *52-53*, 436-446. https://doi.org/10.1016/j.ymssp.2014.08.007

[14] Chen, J., Zhang, L., Li, Y., Shi, Y., Gao, X., & Hu, Y. (2022). A review of computing-based automated fault detection and diagnosis of heating, ventilation and air conditioning systems. *Renewable and Sustainable Energy Reviews*, *161*, 112395. https://doi.org/10.1016/j.rser.2022.112395

[15] Hashim, H., Ryan, P., & Clifford, E. (2020). A statistically based fault detection and diagnosis approach for non-residential building water distribution systems. *Advanced Engineering Informatics*, *46*, 101187. https://doi.org/10.1016/j.aei.2020.101187

[16] Liu, J., Guo, X., & Deng, L.J. (2018) Resistance variant single fault source diagnosis of mine ventilation system based on air volume characteristic. *Journal of China Coal Society*, *43*(1), 143-149.

[17] Liu, J., Liu, L., & Huang, D. (2020) Resistance variant fault diagnosis of ventilation system based on composite features of air volume and air pressure. *Journal of Safety Science and Technology*, *16*(1), 85-91.

[18] Huang, D., Liu, J., & Deng, L. (2019). An adaptive Kalman filter for online monitoring of mine wind speed. *Archives of Mining Sciences*, *64*(4), 813-827. https://doi.org/10.24425/ams.2019.131068

[19] Zhou, Q.C., Liu, J., & Liu, L. (2019). Research on fault diagnosis penalty coefficient and kernel function coefficient optimization of ventilation system based on SVM. *Journal of Safety Science and Technology*, *15*(4), 45-51.

[20] Ni, J.F., Li, Z., & Le, X.R. (2022). Resistance variant fault diagnosis method of ventilation network based on random forest. *Journal of Safety Science and Technology*, *18*(4), 34-39.

[21] Ni, J.F., Le, X.R., & Chang, L.F. (2021). Resistance variant fault diagnosis and optimized layout of sensors for mine ventilation based on decision tree. *Journal of Safety Science and Technology*, *17*(2), 34-39.

[22] Zhang, L., Zhang, Y.H., & Zhang, Y.B. (2022). Research on fault diagnosis method of ventilation network based on machine learning. *Journal of Mine Automation*, *48*(3), 91-98.

[23] Zhao, D., & Shen, Z. (2022). Study on Roadway Fault Diagnosis of the mine ventilation system based on improved SVM. *Mining, Metallurgy & Exploration*, *39*(3), 983-992. https://doi.org/10.1007/s42461-022-00595-z

[24] Wang, D., Liu, J., Deng, L., & Wang, H. (2022). Intelligent diagnosis of resistance variant multiple fault locations of mine ventilation system based on ML-KNN. *PLOS ONE*, *17*(9), e0275437. https://doi.org/10.1371/journal.pone.0275437

[25] Liu, L., Liu, J., Zhou, Q., & Huang, D. (2022). Machine learning algorithm selection for windage alteration fault diagnosis of mine ventilation system. *Advanced Engineering Informatics*, *53*, 101666. https://doi.org/10.1016/j.aei.2022.101666

[26] Liu, L., Liu, J., Zhou, Q., & Huang, D. (2023). Influence of sample attributes on generalization performance of machine learning models for windage alteration fault diagnosis of the mine ventilation system. *Expert Systems with Applications*, *213*, 119320. https://doi.org/10.1016/j.eswa.2022.119320

[27] Creswell, A., & Bharath, A.A. (2019). Inverting the generator of a generative adversarial network. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(7), 1967-1974. https://doi.org/10.1109/tnnls.2018.2875194

[28] Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein GAN adversarial networks. *Proceedings of the 34th International Conference on Machine Learning (ICML)*, *70*, 214-223.

[29] Gulrajani, I., Ahmed, F., & Arjosky, M. (2017). Improved training of wasserstein GANs. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 5767-5777.

[30] Wu, J., Huang, Z., & Thoma, J. (2018). Wasserstein divergence for GANs. *Proceedings of the 15th European Conference on Computer Vision*, 653-668. https://doi.org/10.1007/978-3-030-01228-1_40

[31] Zhao, D., Shen, Z.Y., & Song, Z.H. (2023). Intelligent fault diagnosis of mine ventilation system for imbalanced data sets. *Journal of China Coal Society*, *48*(11), 4112-4123. https://doi.org/10.13225/j.cnki.jccs.2022.1872

Модель машинного навчання для діагностики відхилень опору у вентиляційній системі шахти за умов незбалансованих даних

З. Шен, М. Ян, Д. Зао

Мета. Підвищення точності діагностики несправностей, пов'язаних зі зміною аеродинамічного опору у вентиляційних системах шахт, за умов незбалансованих даних на розробки інтегрованої методики, що поєднує генеративне балансування вибірки та ансамблеву глибоку класифікацію.

Методика. На основі WGAN-div виконується штучне розширення незбалансованих даних з метою ефективного збільшення початкового масиву. Для забезпечення діагностики несправностей у вентиляційній системі інтегровано ансамблевий метод Bagging та глибоку нейронну мережу ResNet.

Результати. Сформовано експериментальні набори даних з коефіцієнтами незбалансованості 1:2, 1:8, 1:10 та 1:20 на прикладі простої Т-подібної вентиляційної мережі. Детально проаналізовано вплив незбалансованих даних на діагностику відхилень опору (WAFs) у вентиляційній системі. Виконано порівняльні експерименти із застосуванням різних моделей збільшення даних та моделей класифікації для вентиляційної системи шахти Донгшань Для оцінювання ефективності моделей використано низку показників, а також візуалізацію t-SNE. Результати показують, що дані, згенеровані моделлю WGAN-div, з вискоим рівнем достовірності узгоджуються з реальними, при цьому у порівнянні з GAN, WGAN і WGAN-GP модель WGAN-div демонструє кращі характеристики, а продуктивність глибокої моделі ResNet також суттєво зросла.

Наукова новизна. У роботі вперше проведено дослідження діагностики несправностей вентиляційних систем за умов незбалансованих даних як на рівні самих даних, так і на рівні мережевої моделі, що дозволяє ефективно розв'язати проблему дисбалансу, характерну для реальних умов роботи вентиляційних систем шахт.

Практична значимість. Запропонований підхід може слугувати технічною основою для впровадження інтелектуальних систем вентиляції, підвищуючи надійність моніторингу та загальний рівень безпеки вентиляційних систем шахт.

Ключові слова: *вентиляційна система шахти, діагностика несправностей, незбалансовані дані, генеративно-змагальна мережа, Bagging-ResNet*

Publisher's note