WEP
Warwick Evans Publishing

Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 7
TPCEE 2024

# Prediction of Electric Load Neural Network Prediction Model for Big Data

Jialin Li*, Binbin Li#, Zheng Luo#, Weikang Chen#

School of Faculty of Financial Mathematics and Statistics, Guangdong Finance University, Guangzhou, China, 510521

* Corresponding Author Email: 13713512595@163.com

#These authors contributed equally.

**Abstract.** Rock bursts are one of the most common and hazardous disasters encountered in coal mining operations, particularly in China. The occurrence of rock bursts poses significant risks to the safety of miners and the structural integrity of mining operations. Consequently, coal mines are required to conduct extensive and frequent rock burst risk assessments, which contribute to the increasing workload and operational pressures on mine personnel. Addressing this challenge is critical for enhancing the safety and efficiency of mining activities. In response to this issue, this paper proposes a novel approach to predicting rock bursts by preprocessing electromagnetic radiation (EMR) and acoustic emission (AE) signals to remove noise and interfering signals. The preprocessing step is crucial as it helps in isolating the relevant features that are indicative of potential rock bursts. By extracting both the interference signal features and the signal precursor features, a comprehensive training set for the predictive model is established. To further enhance the accuracy of predictions, the study employs advanced machine learning techniques, specifically random forests and decision trees, to train and analyze the signal features. These models are chosen for their robustness and ability to handle complex data patterns, making them well-suited for the prediction task. The experimental results are promising, demonstrating that the model achieves a high prediction accuracy of 99.39%, with an Area Under the Curve (AUC) value of 0.99, indicating excellent performance. This high level of accuracy suggests that the model is highly effective in predicting rock bursts, thereby offering a valuable tool for improving mine safety and reducing the workload associated with manual risk assessments.

**Keywords:** Data Preprocessing; Fourier; Random Forest; Sliding Window; Decision Tree.

## 1. Introduction

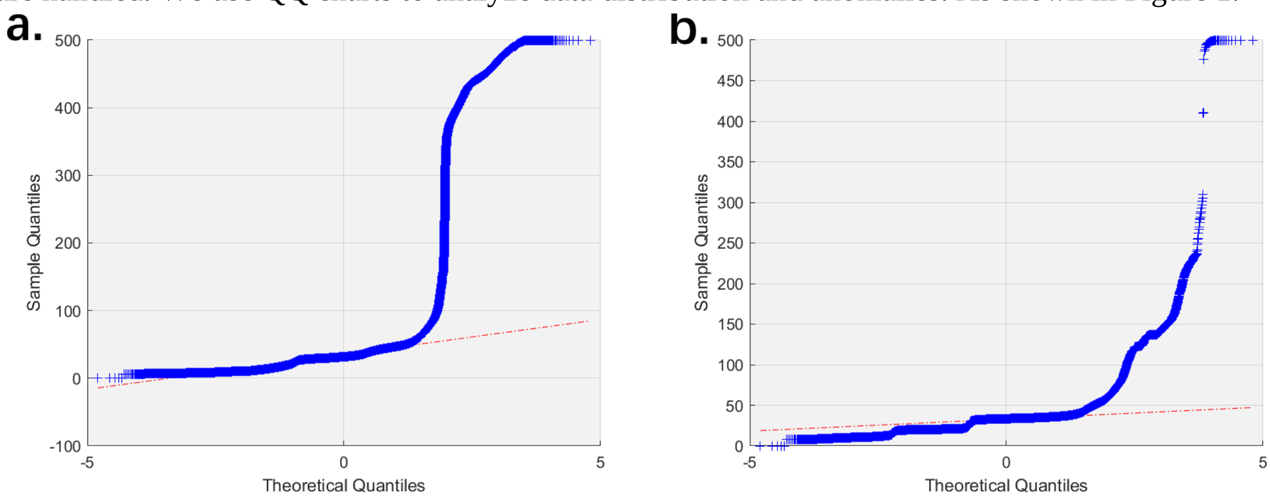
Rock burst is a typical coal-rock dynamic disaster in coal mining engineering, and with the increase of coal mining depth, rock burst accidents occur frequently, which seriously threatens mine safety and underground personnel safety [1-3]. Predicting the danger of rock bursts in deep coal mining is of great significance to the personal safety and economy of the people.

To monitor the rock burst problem of coal mines, the commonly used methods include the comprehensive index method, the multi-factor pattern recognition method, the geological structure prediction, the mining settlement prediction, the mining depth prediction, and the coal-rock impact tendency prediction. The commonly used methods for local prediction are rock mechanics and geophysical methods, such as microseismic monitoring, seismic CT, charge method, electromagnetic radiation monitoring, roof stress monitoring,etc[4-9].

To solve the above problems, this paper firstly uses a QQ map to preprocess the rock burst dataset, then extracts the features of the interference signal through fast Fourier, and extracts the features of the precursor feature data by sliding window, to form the model training set and the test set, then carries out random sample oversampling and data standardization, and finally selects the decision tree model for classification prediction.

## 2. Data Source and Quality Check

The data for this study is derived from https://51mcm.cumt.edu.cn/, and the data contains signal intensity and time information for electromagnetic radiation (EMR) and acoustic emission (AE) over multiple periods. We know that there are a large number of interference signals in some electromagnetic radiation and acoustic emission signals of the working face on the site, which may be caused by other operations or equipment interference on the working face, which has a certain impact on the processing of electromagnetic radiation and acoustic emission signals in the later stage. Therefore, to maintain the safety of coal mines, we will extract the characteristics of interference signals. First of all, to improve the quality of the data, it is cleaned and the missing values and outliers are handled. We use QQ charts to analyze data distribution and anomalies. As shown in Figure 1.




(a) EMR QQ plot (b)AE QQ plot

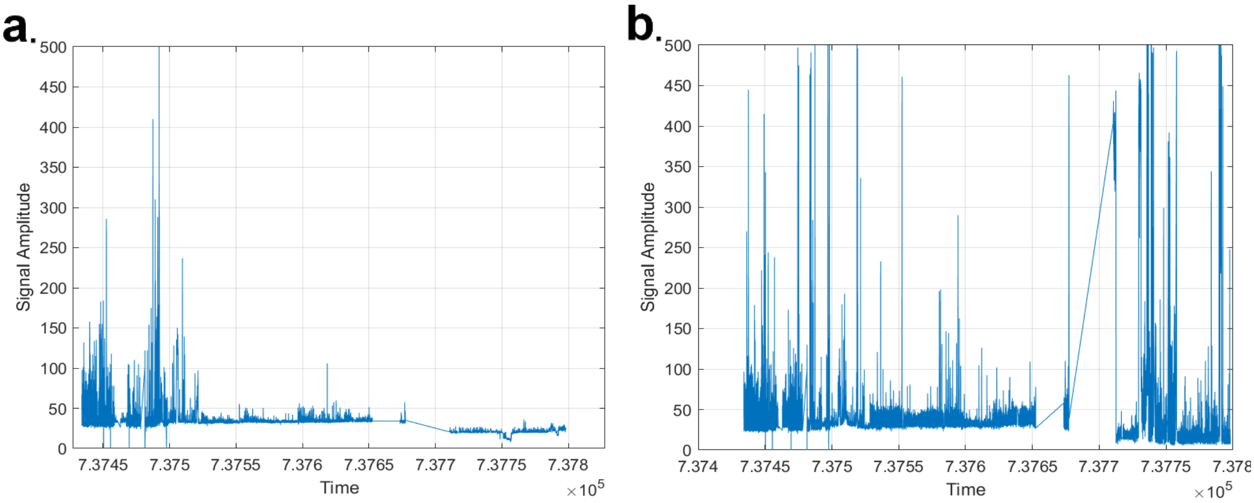
**Figure 1.** QQ plot

The data obtained from the analysis of the legend deviates from the normal distribution as a whole, and there are many anomalies. However, we cannot directly judge the processing of values from the data, there are many different signal data at the same time, and to make full use of each data, we default to this situation because multiple sensors are recording at the same time [10]. Therefore, in the following, we will extract features and related judgment operations, and only use the averaged features. Instead of using the features of the summation. As shown in Table 1.

**Table 1.** Examples of Different Data at the Same Time

| EMR | time | class |
|---|---|---|
| 52.3 | 2019-01-09 04:59:24 | D/E |
| 46 | 2019-01-09 04:59:24 | D/E |

Next, we filter the interfering signal data (Category C). As shown in Figure 2.




(a) AE Interference Signa　　　　(b) EMR Interference Signal

**Figure 2.** Interference Signal

There is no fixed pattern of interference signal data, its intensity fluctuates in the range of 100-500, and it has no fixed period and interval, the randomness of this intensity, may be due to the superposition effect of a variety of interference sources in the process of going deep underground, each interference source is independent of each other, affecting the data in its specific way, thus producing different interference data signals.

In addition, the identification of outliers is particularly important in dynamic data, through statistical regression analysis of data such as mean, median, and mode variance, the characteristics of the data are not found for the time being, but the nature of their fluctuations can be understood more deeply. No cyclical patterns have been found, but there may be other data patterns on longer time scales.

## 3. Feature Extraction and Frequency Domain Analysis

### 3.1. To construct external signal features, firstly, the interfering signal is extracted in the time domain.

Kurtosis indicates how sharp the data distribution is

$$Kurtosis = \frac{N(N+1)}{(N-1)(N-2)(N-3)} \sum_{i=1}^{N} (\frac{x_i - \mu}{\sigma})^4 - \frac{3(N-1)^2}{(N-2)(N-3)} \tag{1}$$

Pulse Count is typically used to measure the number of pulses in a signal that exceeds a certain threshold.

$$Pul\,se\,Count = \sum i = 1^N 1(x_i > T) \tag{2}$$

### 3.2. Time-domain feature extraction of interference signals.

A discrete signal is a function in which the value of a signal is defined only at a series of discrete points in time or space. In the data, electromagnetic radiation and acoustic emission sensors collect data every 30 seconds, and the change trend of these data can be used to determine whether there is a danger of rock burst in the current working face or roadway, so the collected signals are discrete [11].

Through the Fourier transform, the time-domain signal is converted into a frequency-domain representation, so that the periodic patterns, frequency components of the signal, and their relative importance, which are not easy to observe in the time-domain, can be demonstrated in the frequency domain [12]. Next, we use the Fourier transform to extract the interfering signals from the sequence

Spectral energy is the total energy of a signal in the frequency domain and can be calculated by the following formula:

$$Spectral\ Energy = \sum_{k=0}^{N-1} \left| X[k] \right|^2 \tag{3}$$

The spectral Peak is the value with the largest amplitude in the signal spectrum and can be calculated by the following formula

$$Spectral\ Peak = \max\left( \left| X\left[k\right] \right| \right) \tag{4}$$

Spectral Density represents the average energy density of the signal spectrum and can be calculated by the following formula:

$$Spectral\ \ Density = \frac{1}{N} \sum_{k=0}^{N-1} \left( \frac{\left| X\left[k\right] \right|^2}{N} \right) \tag{5}$$

The Discrete Fourier Transform (DFT) is a method for transforming a finitely long discrete sequence into the frequency domain.

$$X\left[k\right] = \sum_{n=0}^{N-1} x\left[n\right] \cdot e^{-j\frac{2\pi}{N}kn} \tag{6}$$

x[n] is the n-th sample of the time-domain signal, N is the length of the signal, k is the frequency index, and i is the imaginary unit.

As shown in Table 2.

**Table 2.** Extraction of Time Domain and Frequency Domain Features

| Features | EMR | AE |
|---|---|---|
| average value | 77.9585 | 59.9691 |
| variance | 8229.3701 | 6467.6703 |
| kurtosis | 12.9982 | 26.0055 |
| Spectral energy | 75231750 | 12965982 |
| The main frequency component | 0.0001902 | 0.0007758 |
| Peak spectrum | 409983.77 | 77300.177 |
| Spectral density | 14.3053 | 10.0589 |
| Pulse counting | 128 | 13 |

The original time-domain signal is properly pre-processed to reduce noise and improve the quality of the signal. Then, the Fast Fourier Transform (FFT) algorithm is used to convert the signal from the time domain to the frequency domain to generate a spectrogram, which shows the amplitude of different frequency components and provides a more comprehensive understanding of the frequency domain characteristics of the data [13]. As shown in Figure 3 and 4.

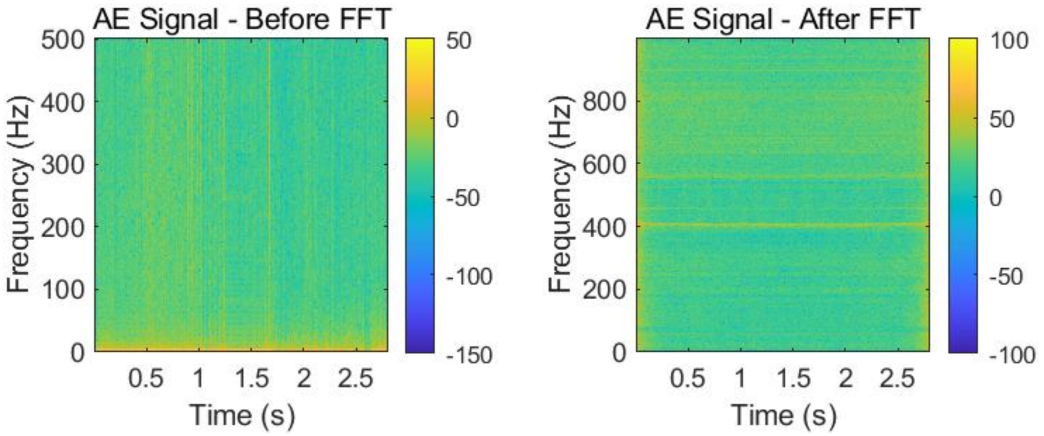


**Figure 3.** Comparison of AE Interference Signals Before and After the FFT plot;

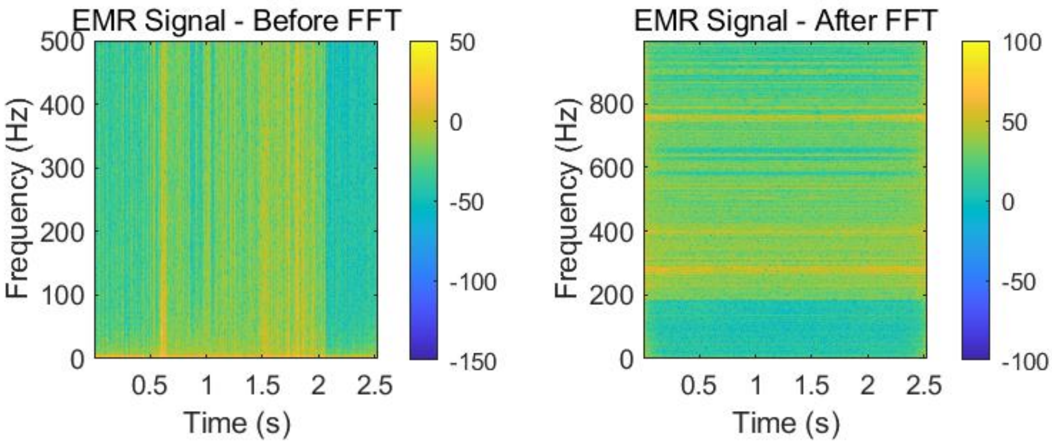


**Figure 4.** Comparison of EMR Interference Signals Before and After FFT plot

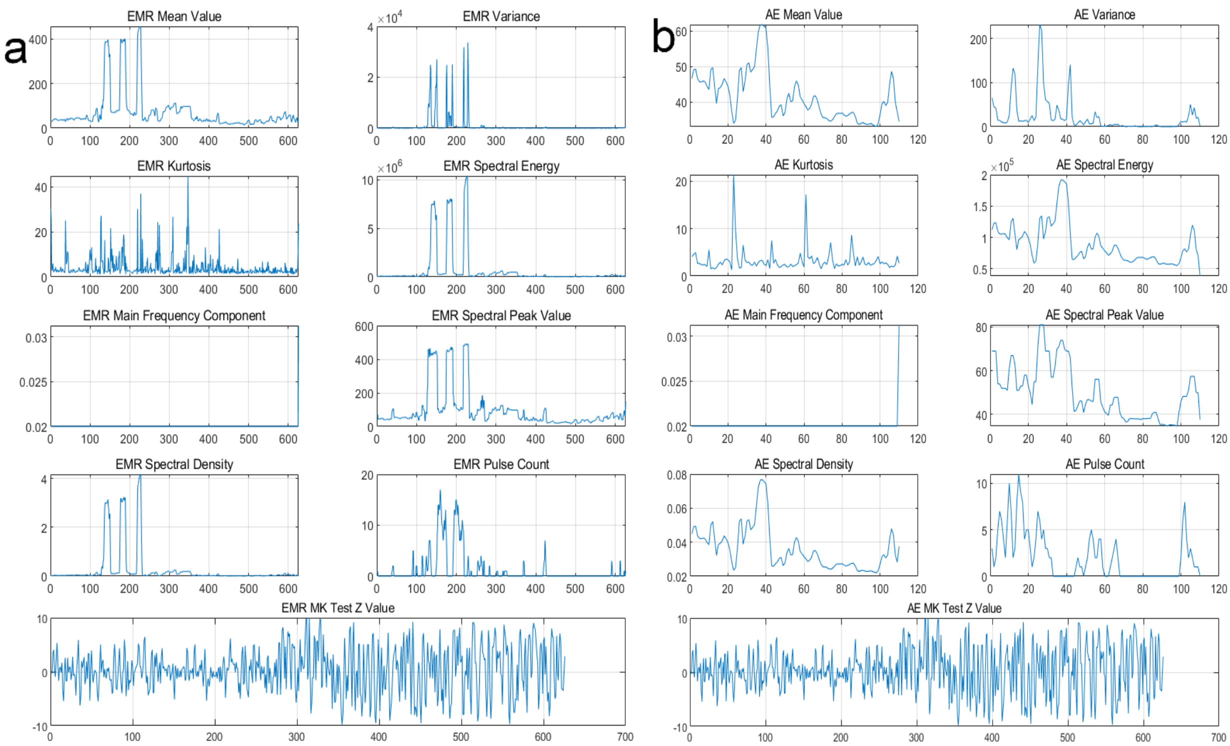
## 4. Slide the window

To identify the signal data in the dataset, the characteristics of the time-varying signal can be captured by using a sliding window to identify the features of the dataset, which can adapt to the local changes of the signal, reduce the amount of computation, and provide real-time or near-real-time analysis results [14].

Since the feature data of each time point do not exist separately, the data before and after the time node are analyzed together and then the feature extraction is carried out, so the method of selecting the sliding window can be used to extract the precursor features of electromagnetic radiation and acoustic emission signals. The extracted data calculates the 6 characteristics of mean, standard deviation, peak, pulse count, major frequency components, and spectral energy. Select the data window size of 50 for the data in the dataset, and set the error rate of the extracted feature text to 0.1

$$
judge\ \text{model}=\begin{cases} 1(\text{in the scrambling signal range}), \\ if\,(any(x_i<(1+rate)\cdot feature_i \text{ and } \mathrm{x}_i>(1-rate)\cdot feature_i)) \\ 0(not \text{ in the scrambling signal range}), \\ if\,(\text{not } any(x_i<(1+rate)\cdot feature_i \text{ and } \mathrm{x}_i>(1-rate)\cdot feature_i)) \end{cases} \tag{7}
$$

xi I represents the signal characteristics of each group of interference signals represents features of the extracted feature results, and the rate represents the set error rate. The results are visualized after sliding window feature extraction of the data in the dataset. As shown in Figure 5.




(a) Feature extraction of E MR (b) Feature extraction of AE.

**Figure 5.** Feature extraction, Each of the remaining subgraphs represents a feature in the title

## 5. Performance indicators of random forests:

When applying random forest models for signal recognition, the performance indicators of the model need to be evaluated to ensure its accuracy and reliability [15]. The following are the main performance indicators of the random forest model in the identification of electromagnetic radiation (EMR) and acoustic emission (AE) signals:

Accuracy: Describes the proportion of correctly predicted (positive and negative) by the model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

Precision: Describes the proportion of samples predicted to be positive that are positive. This is an important metric to measure the accuracy of the model in predicting positive class samples.

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

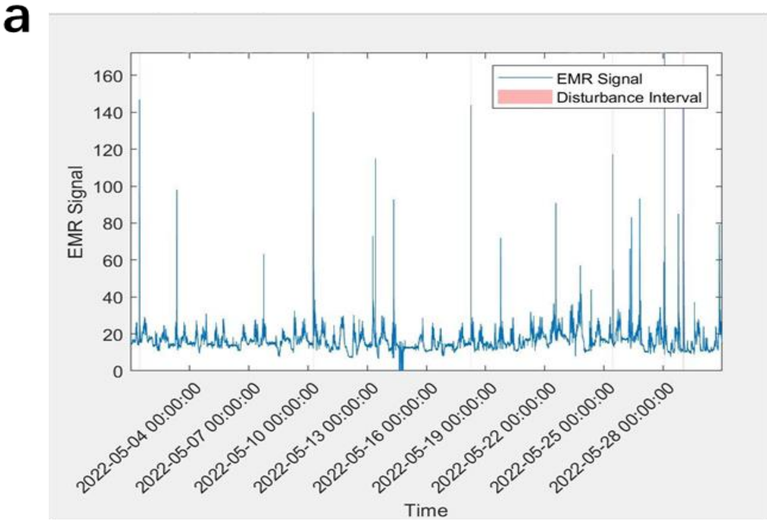
Recall: Describes the proportion of all actual positive samples that were correctly predicted by the model to be positive. This is an important metric for measuring the model's ability to capture positive class samples.

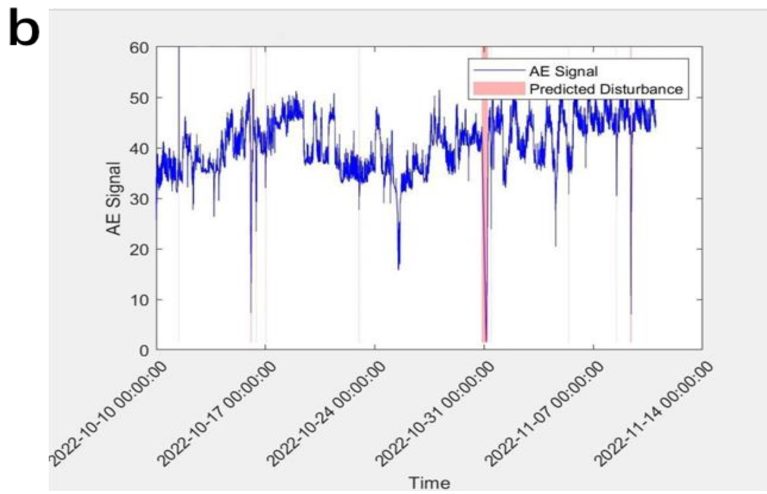
$$Recall = \frac{TP}{TP + FN} \tag{10}$$

F1 Score: The F1 score is a blended average of precision and recall and is a combination of these two metrics.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{11}$$

The electromagnetic radiation and acoustic emission data in a specific period are imported into the model for classification and judgment. The random forest model realizes high-precision identification of interference signals. The specific results are as follows Figure 6 and Table 3.




(a) Interference interval signal of EMR; (b)Interference interval signal of AE

**Figure 6.** Interference interval signal.

**Table 3.** Performance index of random forest model

| Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|
| 99.80% | 1.00 | 0.97 | 0.98 |

The results show that the random forest model has high accuracy and reliability in identifying the interference signal and the precursor characteristic signal in the signal, which provides important technical support for the prediction of rock burst hazards in deep coal mining.

Through the MATLAB decision tree model, the model is trained on the training set of electromagnetic radiation and acoustic emission signals, and the solution results of the decision tree are trained by dividing the training proportion of the dataset, to test whether the results of the trained decision tree solution are correct due to the remaining part of the data. The specific results are as follows Figure 7 and Figure 8.

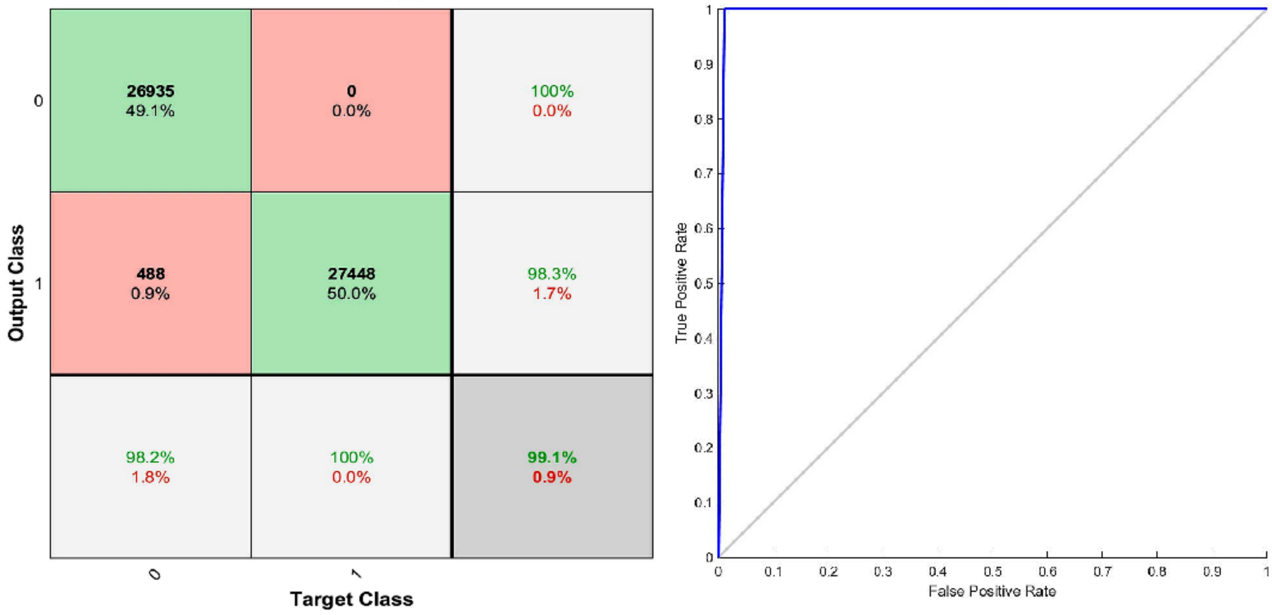



**Figure 7.** EMR confusion matrix and ROC curve

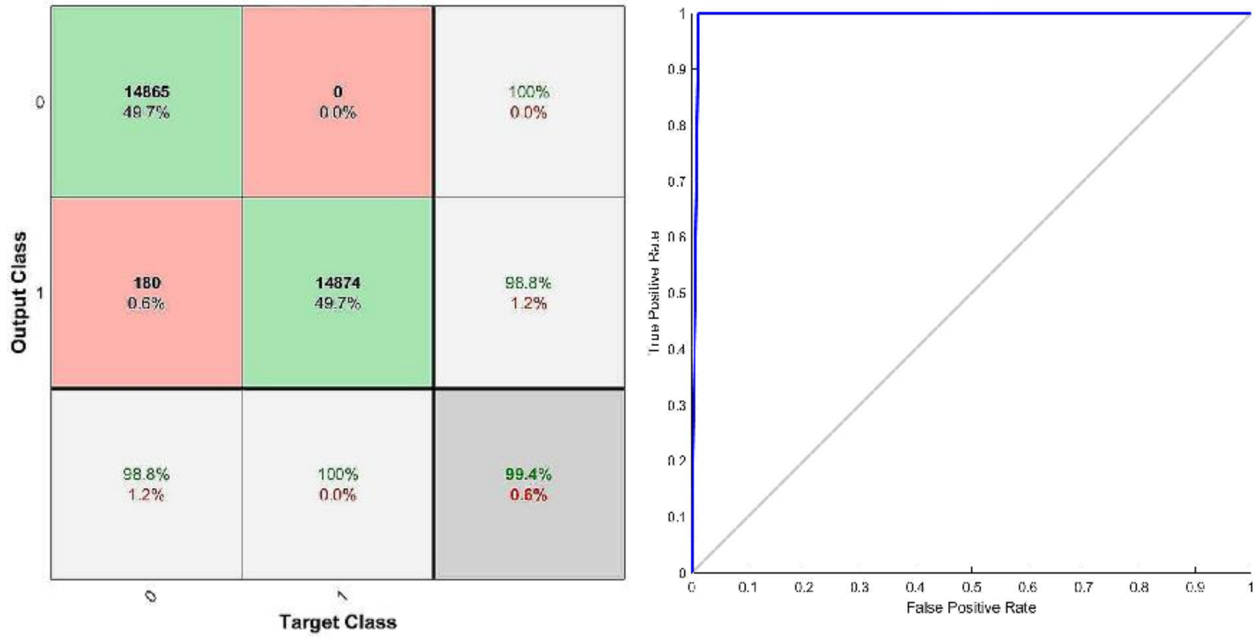



**Figure 8.** AE confusion matrix and ROC curve

The analysis of the confusion matrix and ROC curve can be used to evaluate the classification performance in more detail, including the accuracy and recall of different categories, and the overall classification performance of the model under different thresholds can be more intuitively demonstrated [16]. The optimal classification model can be selected, the model parameters can be adjusted, and the generalization ability of the model can be evaluated. The confusion matrices of the two models show that they have very high accuracy, precision, and recall in signal recognition, indicating that they have good classification performance. As shown in Table 4 and Table 5.

**Table 4.** Performance of EMR decision tree model

| Accuracy | Precision | Recall | F1 Score | AUC |
|---|---|---|---|---|
| 99.50% | 0.99 | 1.00 | 1.00 | 0.99 |

**Table 5.** AE decision tree model performance indicators

| Accuracy | Precision | Recall | F1 Score | AUC |
|---|---|---|---|---|
| 99.39% | 0.99 | 1.00 | 0.99 | 0.99 |

Through feature extraction and model training of electromagnetic radiation and acoustic emission signals, the decision tree model can effectively identify the features in the signals and provide technical support for related applications [17].

## 6. Conclusions

This paper tackles the problem of signal interference encountered during the monitoring of electromagnetic radiation and acoustic emission signals. It examines the preprocessing techniques from both the characteristics of interference signals and the features of precursor signals to select a training dataset for machine learning models. The application centers on random forest and decision tree approaches for predicting shock ground pressure. Experimental outcomes reveal that the decision tree model attains a prediction accuracy of 99.39% for shock ground pressure, signifying its efficacy in prediction.

Additionally, to enhance the prediction accuracy in diverse coal mining circumstances, it is requisite to re-train and adjust the model in accordance with the differentiated actual signal data to boost the prediction accuracy. By using the decision tree model, we believe that the occurrence of rock burst can be measured more accurately in the future prediction, and the prevention and control of rock burst can play a better effect, so that the safety of miners can be better guaranteed.

## References


[1] JIANG Fuxing, ZHANG Xiang, ZHU Sitao. Discussion on key problems in prevention and control system of coal mine rock burst [J]. Coal Science and Technology,2023,51(1):203−213.

[2] ZHU Quanjie, GU L ei, CHENG Yunhai, et al. Design and implementation of evaluation software for coal mine rockburst prevention system [J]. Journal of North China Institute of Science and Technology, 2022,19(05):1-7.

[3] QI Qingxin, LI Yizhe, ZHAO Shankun, et al. Seventy years development of coal mine rockburst in China: Establishment and consideration of theory and technology system[J]. Coal Science and Technology,2019,47(9):1-40.

[4] WANG Chao, ZHANG Guangchao, ZHAO Xipo, et al. Evaluation method and application of impulse hazard based on roof structure characteristics [J]. Journal of Mining and Strata ControlEngineering,2024,6(2):125-135.

[5] LAN T W, ZHANG Z J, SUN J W, et al. Regional prediction and prevention analysis of rockburst hazard based on the Gaussian process for binary classification [J]. Frontiers in Earth Science,2022,10:959232.

[6] LIU Yaoqi, CAO Anye, WANG Yao, et al. Prediction method of coal burst based on attenuation characteristic seismic cluster energy[J]. JOURNAL CHINA COAL SOCIETY,2022,47(4):1523-1533.

[7] LIU Yaoqi,CAO Anye, WANG Songwei,et al. Prediction method of coal burst based on attenuation characteristics of seismic clusterenergy [J]. Journal of China Coal Society, 2022, 47(4): 1523 -1533.

[8] YAO Jingming, XU Ziwen, WANG Jian, et al. Prediction of rockburst electromagnetic radiation based on multifractal theory [ J ]. Mining Safety & Environmental Protection,2021,48(5):59-63.

[9] ZHAO T B, ZHANG P F, GUO W Y, et al. Controlling roof with potential rockburst risk through different pre-crack length: Mechanism and effect research [J]. Journal of Central SouthUniversity,2022,29(11):3706-3719.

[10] WENTingxin, LIYangzi. Risk prediction model of rock burst based on preprocessing for AFOA-ELM [J]. China Safety Science Journal, 2019,29(08):

[11] Wang Zhonghao. Stress Monitoring and Risk Prediction of Rock Burst in Mianhuakeng Mine [D]. China University of Geosciences (Beijing), 2020.

[12] Qu Hongquan, Wang Zhengyi, Sheng Zhiyong, Qu Hongbin, Wang Ling. Based on Gradient Boosting Decision Tree Algorithm [J]. Fiber Intrusion Signal Classification, 2022,59(23): 106-113.

[13] ZHENG Liang, LIU Ning . Identification of the Formation Method of Toner-Base Printed; Images Using Two-dimensional Fast t Fourier Transform [J/OL] . Forensic Science and Technology, 2024.

[14] BAI Xiao pin, LIU Yuhang. Life Analysis Rolling Bearing Based on K-means Degradation Identification and Random Forest [J]. Modular Machine Tool & AutomaticManufacturingTechnique,2024, (07):150-155+160.

[15] NIU Rongze, ZHANG Kai, XIE Ruirui, et al. Research on monitoring system of cable network pipeline based on random forest [J/OL]. Electrical Measurement & Instrumentatio,1-9[2024-07-21].

[16] Wu Bingmei. Study on prediction model of rockburst hazard based on supervised learning [D]. Anhui University of Science and Technology,2023.

[17] SU Guorui. Research on Coal Mine Accidents Key Hidden Dangers Mining and Safety Risk Prediction [D]. Liaoning Technical University,2022.