

¹ Enke Hou^{2,*} Mengfan Yang³ Bingbing Xia

Microseismic B-Value based Time Series Prediction of Rock Bursts



Abstract: - Monitoring the microseismic activity in mines helps to understand the interference range and characteristics of the microseismic activity triggered by mining activities, thereby predicting the future trend of rock burst pressure changes based on temporal characteristics. This paper collects microseismic monitoring data and applies the G-R relationship to calculate the microseismic b-value, finding that before the occurrence of rock burst events, the b-value is abnormally low, falling below the average value, which is used as a predictive indicator. Using a deep learning algorithm combining Complementary Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), Northern Goshawk Optimization (NGO), Convolutional Neural Network-Gated Recurrent Unit (CNN-GRU), and Self-Attention mechanism, a rock burst prediction model is constructed for advanced prediction of the b-value. Compared with different algorithm models, it has a better predictive effect. The model can effectively learn from the characteristic information of the historical microseismic b-value and make predictions. The accuracy of the advanced single-step prediction is quite high without any obvious lag signs, providing a basis for early warning of rock burst events.

Keywords: Impact Earth Pressure, Microseismic b-value, Deep Learning, Time Series, Microseismic Early Warning.

INTRODUCTION

Following extensive and prolonged mining operations, China's shallow coal resources are dwindling rapidly. The focus of production is inching towards deeper strata with an incremental escalation in mining intensity ^[1]. The sporadic nature of rock bursts presents a considerable challenge to mine safety. To effectively prevent and reduce the damage caused by rock bursts, it is imperative to thoroughly comprehend the underlying mechanisms and develop proficient early warning systems for prediction and prevention ^[2]. Variations in microseismic parameters can serve as predictive indicators for geopressure disasters ^[3].

Empirical formula-based studies primarily leverage historical data and on-site experience to establish predictive models. Their limitation is that a single empirical formula often struggles to accommodate the intricate complexities presented by diverse mining conditions ^[4].

Statistical methods analyze patterns within historical data to anticipate the likelihood of rock bursts. These methods are constrained by their dependence on the quantity and quality of data available and are often unequipped to manage nonlinear relationships ^[5].

Through meticulous investigation into the mechanisms behind rock bursts, including microseismic activity patterns and regional traits, researchers like Yuan Ruifu have identified precursory indicators that precede rock bursts. These encompass periods of heightened microseismic activity, brief calm intervals, and a predilection for occurrences in specific zones. These findings pave the way for theoretical model construction and lay the groundwork for subsequent data mining techniques, thereby augmenting the capability to discern and forecast impending risks in mining regions using machine learning algorithms ^[6].

With advancements in technology, research anchored in machine learning, particularly deep learning, is gaining

¹ Enke Hou, Affiliation: School of Geology and Environment, Xi'an University of Science and Technology, Xi'an Shaanxi, 710054, China.

^{2*} Corresponding author: Mengfan Yang, Affiliation: School of Geology and Environment, Xi'an University of Science and Technology, Xi'an Shaanxi, 710054, China. Email:1686667533@qq.com

³ Bingbing Xia, Affiliation: Shaanxi Coalfield Geophysical Surveying and Mapping Co., Ltd, Xi'an Shaanxi, 710005, China.

momentum. Machine learning, especially deep learning methodologies, discern patterns in rock burst incidents by training on extensive datasets. Despite demonstrating commendable predictive capabilities, these methods face limitations in feature selection, model development, and dataset curation^[7]. In particular, selecting features that are closely associated with rock bursts remains a challenge. Furthermore, crafting deep learning models that authentically reflect the dynamic nature of rock bursts is crucial^[8]. Additionally, the scarcity of large-scale, high-quality datasets impedes the enhancement of deep learning model performance.

Numerous scholars employ various techniques to monitor rock bursts, such as assessing the stress levels in rocks, analyzing the energy and frequency of microseismic events, and evaluating the acoustic emission properties of rocks^[9-14]. Domestic and international researchers have proposed numerous precursory indicators based on diverse monitoring techniques, including the frequency of mine tremors, seismic energy, b-value, missing earthquakes, total fault area, and concentration of seismic sources^[15-19]. These indicators serve as advance warning signs, offering multifaceted perspectives on the stability of rock masses and the potential for rock bursts. Identifying precursory information during the gestation and occurrence of rock bursts, along with monitoring the dynamic fluctuations of various indicators, facilitates precise predictions and timely early warnings for rock burst disasters. Nevertheless, the correlation between rock bursts is not linear; they exhibit elements of fuzziness and uncertainty. Concurrently, the complexity of geological conditions can impact the effectiveness of these indicators across different mining environments^[20].

This research is conducted against the backdrop of Mengcun Coal Mine and focuses on the microseismic monitoring data collected as the subject of study. The aim is to develop an intelligent early warning model by analyzing the patterns of change in microseismic activity parameters prior to rock failure. This approach seeks to enhance the automation and predictive capabilities of microseismic monitoring systems within mines.

I. RESEARCH METHODOLOGY

A. Research Background

Under specific mining technological conditions, there exists a critical depth below which rock bursts are likely to occur. Due to the varying geological and mining technological conditions across different mining regions, the minimum critical depth at which rock bursts can happen significantly differs. For instance, in the Yanzhou and Lu'an mining areas, the critical depth for rock bursts is roughly between 400-500 meters, while in the Shenhua Xinjiang and Pingzhuang mining regions, it ranges from about 150-300 meters.

Table 1. Critical depth for rock burst in some coal mines in China

Mining shaft	Mentougou Coal Mine	Tianchi CoalMine	Fushun, Datong mining area	Chengzi Coal Mine	Datai Coal Mine	Tao Zhuang Coal Mine	Fangshan CoalMine	Tangshan Coal Mine
Critical Depth/m	200	240	250~300	330	460	480	520	540

Mengcun Coal Mine is located in the Binchang mining area, where 90% of the area has coal seams buried deeper than 600m, with the maximum burial depth reaching 890m. As shown in Figure 1, this exceeds the generally recognized critical depth for rock bursts. Therefore, the No.4 coal seam fully meets the depth conditions necessary for the occurrence of rock bursts.

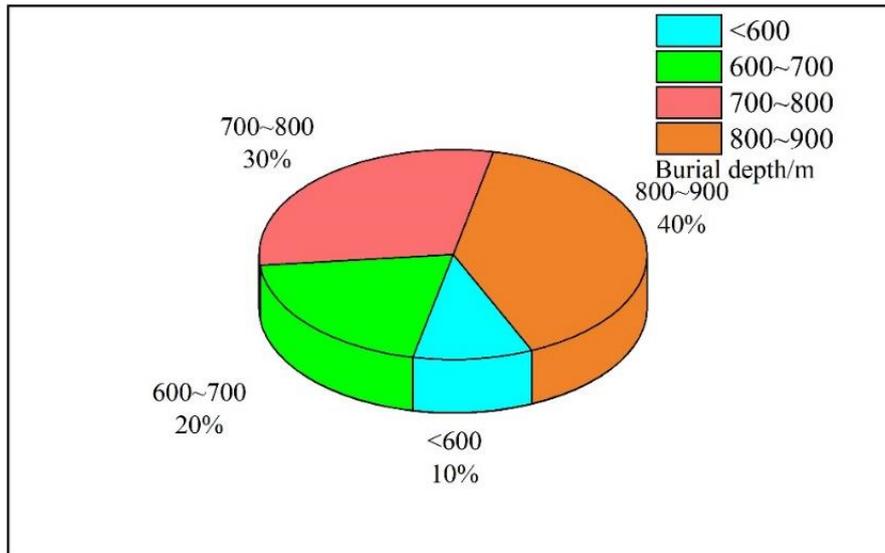


Figure 1 Area proportion of coal seams with different burial depths in Panqu 401 of Mengcun Coal Mine

B. Data Source and Processing

Utilizing the Mengcun SOS microseismic monitoring system to trace and ascertain the microseismic activity at the working face, an investigation into its distribution pattern was conducted. A selection of 31,525 microseismic events recorded in the central gallery from March 2, 2019, to January 10, 2020, served as samples for analysis. As illustrated in Figure 2, the microseismic activity demonstrates systematic characteristics in the spatial distribution of energy.

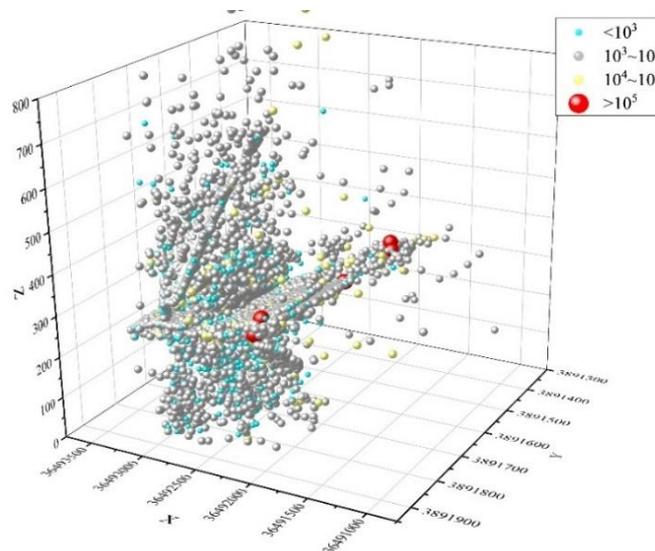


Figure 2 Spatial Distribution Map of Microseismic Events

The analysis of microseismic data has unveiled a significant correlation between the frequency characteristics of the data and the occurrence of bumps. The findings suggest that the frequency of microseismic events tends to mirror the alterations in coal seam structure and its minor ruptures that occur during the mining operation. Conversely, the energy levels often indicate the pressure threshold that the coal mass can withstand, as well as its deformation or fracture strength. As depicted in Figure 3, there is a relationship graph illustrating the interplay between microseismic energy, frequency, and time spanning from March 2, 2019, to January 10, 2020.

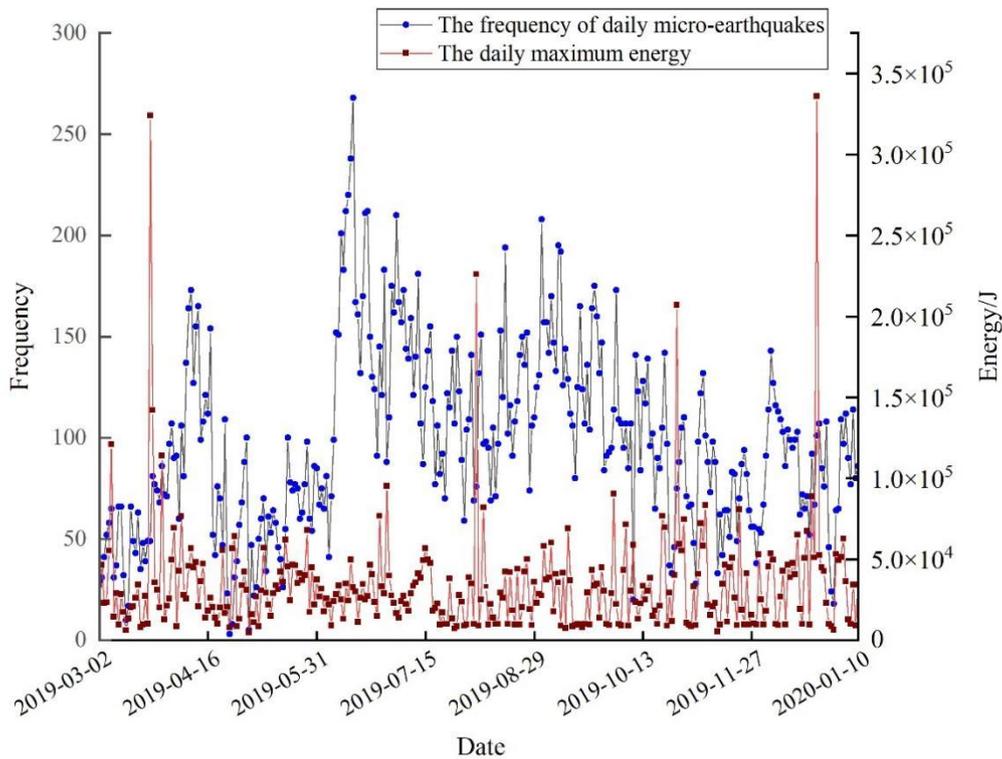


Figure 3 Microseismic Energy, Frequency and Time Relationship Chart

Between March 2, 2019, and January 10, 2020, the vicinity of Zhongyang Alley witnessed a total of seven large-energy microseismic events, each with an energy output exceeding 10^5 J, culminating in a maximum microseismic energy release of approximately 3.36×10^5 J. An examination of the temporal fluctuations in their energy and frequency has led to the conclusion that there is a discernible periodicity in the accumulation and discharge of energy within these substantial microseismic episodes. Drawing from statistical data, a significant dip in frequency correlates with a gradual increase in the daily energy discharged, signaling the imminent arrival of the next bout of energy release. During such intervals, it is prudent to implement necessary measures for shock protection.

In 1914, the seminal G-R relationship was posited by Gutenberg and Richter, establishing the correlation between the frequency of earthquakes and their magnitude. This relationship is enshrined as one of the cornerstone principles in the field of seismology, serving to characterize the distribution of event frequencies across various magnitudes in seismic activity. The formula delineating this relationship is as follows:

$$\lg N = a - bM \quad (1)$$

In the formula, N represents the cumulative number of events with a magnitude greater than M ; M denotes the magnitude of the microseismic event; a , b are constants.

At the mine scale, the energy $\lg E$ of a microseismic event can be substituted for the magnitude M , as illustrated by the subsequent formula:

$$\lg N = a - b \lg E \quad (2)$$

The "b-value" of the G-R relationship can serve as a predictive indicator for hazardous periods of rock bursts, mirroring the stress level within the medium. An inverse relationship exists where a lower b-value signifies a greater proportion of significant ruptures, thereby increasing the likelihood of inducing rock bursts. The most

prevalent technique currently employed for the computation of the b-value is the least squares method, with its calculation formula being specified as follows:

$$b = \frac{\sum_{i=1}^m \lg E_i \sum_{i=1}^m \lg N_i \times \sum_{i=1}^m \lg E_i \lg N_i}{\sum_{i=1}^m \lg^2 E_i \times (\sum_{i=1}^m \lg E_i)^2} \quad (N_i \neq 0, \text{ And not all are 1}) \quad (3)$$

$$b = 0 \quad (N_i = 0, \text{ And all are 1})$$

In the formula, represents the number of magnitude bins; $\lg E_i$ is the microseismic energy level of the i -th bin; N_i is the number of microseismic events with energy in the i -th bin. Based on equation (3), the b value during various time periods at the working face of Mengcun coal mine can be calculated, thus allowing for a more intuitive observation of the variation pattern of the b value.

To reveal the actual relationship between the microseismic energy level and the number of events, Origin's curve-fitting function was utilized to fit a trend change characteristic diagram for the microseismic monitoring data. According to the data analysis results shown in Figure 4, the monitored values of microseismicity exhibit a trend of gradually increasing before decreasing. This is mainly due to the omission of records for small-scale microseismic events at lower energy levels. To ensure accurate and error-free statistical counting of earthquake events at all levels, microseismic events that occurred within an energy level range of 2.8 to 5.2 were selected for further calculation.

Since these microseismic events exhibit a monotonically decreasing distribution within this energy range, the b-value can be determined with greater precision.

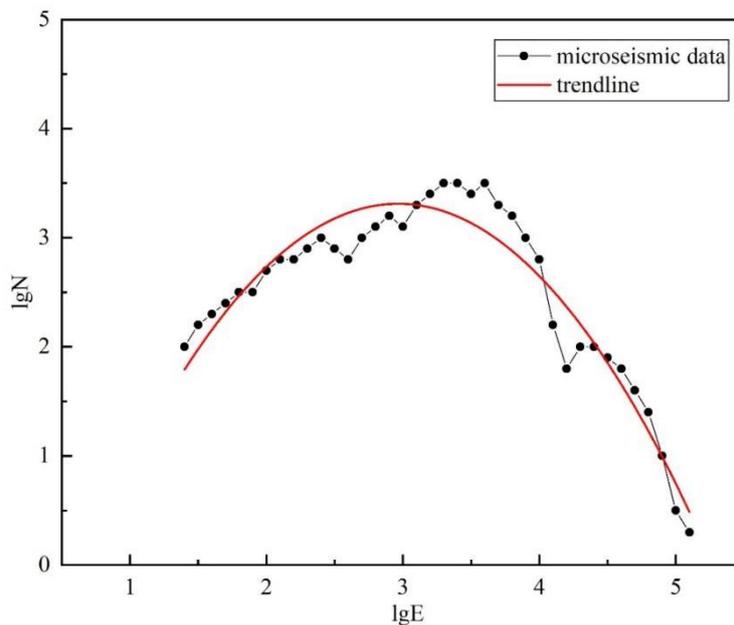


Figure 4 Characteristics of Trend Changes in Microseismic Monitoring Data

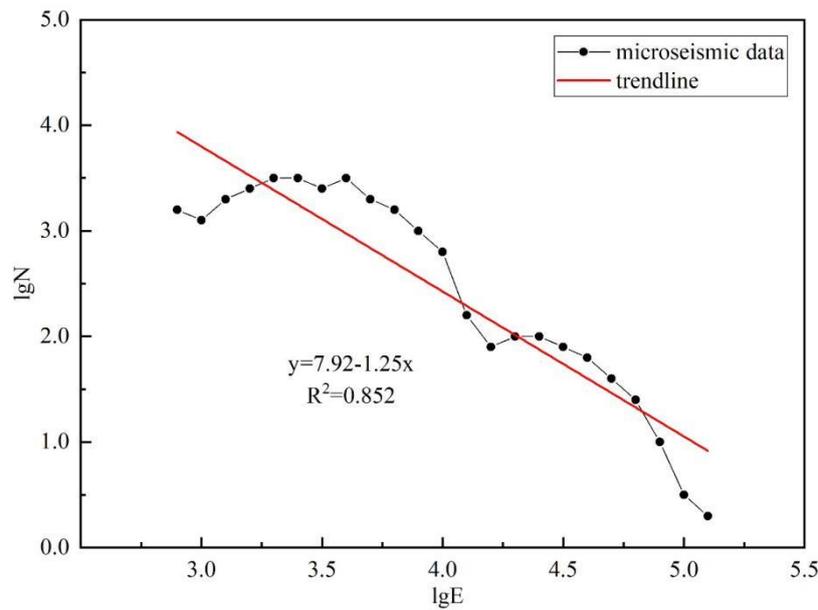


Figure 5 Results of Linear Fitting for Microseismic Monitoring Data

Figure 5 presents the linear fitting outcomes for microseismic monitoring data within a specified interval. The fitting coefficient $R^2=0.852$ can be derived, corroborating the appropriateness of the chosen range of microseismic magnitude levels.

By selecting microseismic energy level samples within the range of 2.8 to 5.2 and employing an increment of 0.1 for the energy level, with a time window set to 5 days, calculations are executed with a time shift interval of 1 day, as depicted in Figure 6. In comparison to the fluctuations observed in other energy and frequency curves within the region, the b-value curve displays more consistent fluctuations and exhibits distinct periodic peak-valley characteristics. High-energy microseismic events predominantly occur in areas with lower b-values, while their frequency of occurrence is relatively diminished in intervals with higher b-values, showcasing a pronounced correlation. Nevertheless, this pattern is not clearly visible in the energy and frequency curves. In instances of large energy level microseismic activities, the corresponding b-values prior and subsequent are typically at reduced levels, particularly when approaching or falling below the mean value of 1.25. This mean value is regarded as the threshold for large energy level microseismic activities. The differentiation of large energy events is thus effective, further substantiating the appropriateness of the selected threshold. Integrated with the actual production circumstances of the mine, this information can serve as a basis for detecting precursors to rock burst incidents. Consequently, the prediction of high-energy microseismic events is advantageous for the proactive prevention of rock bursts, diminishing the likelihood of catastrophe. In the application on-site, intervals with low b-values can be identified as potential periods of impact hazard, during which risk mitigation efforts and personnel management should be intensified to prevent the triggering of rock bursts by high-energy events.

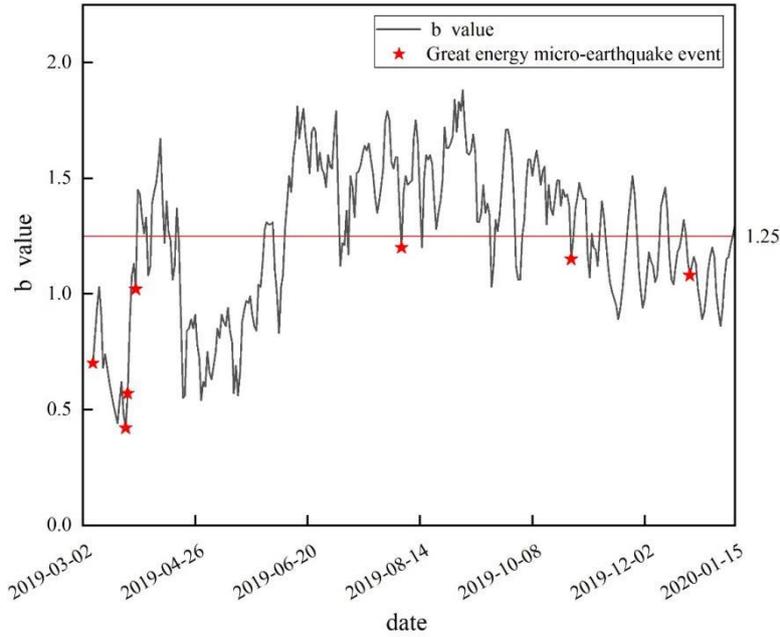


Figure 6 Distribution of b-values and large-energy microseismic events

C. Prediction Model Algorithm Principle

1) Principle of CEEMDAN Algorithm

CEEMDAN is an adaptive posteriori decomposition algorithm, which improves upon the Empirical Mode Decomposition (EMD) and Ensemble EMD (EEMD). It addresses the issues of EMD and EEMD by adding adaptive white noise to the Intrinsic Mode Function (IMF) components of the original data.

The specific steps of CEEMDAN are as follows:

① Adding normally distributed random noise $n(i)$ to multiple data sets x , this process is repeated i times, and the newly obtained sequence is $x(i) = x + \varepsilon_0 n(i)$. Here, ε_0 represents the signal-to-noise ratio of the noise relative to the initial sequence, while $n(i)$ ($i = 1, 2, \dots, N$) denotes the white noise to be added, and i represents the number of times it is added. IMF_i^1 is defined as the modal component obtained after decomposing $x(i)$ using the

EMD method. Correspondingly, for CEEMDAN, the intrinsic modal component IMF_i^1 and the calculated residual component r_i are as follows:

$$IMF_1 = \frac{1}{N} \sum_{i=1}^N IMF_i^1 \quad (4)$$

$$r_1 = x - IMF_1 \quad (5)$$

② $E_k(\mathbb{1})$ is the formula for calculating the k -th order modal component in the EMD decomposition process, and it is used again in the remaining components. White noise is added to the new sequence $r_1(i) = r_1 + \varepsilon_1 E_1(n(i))$, and then EMD decomposition is performed on $r_1(i)$ again to obtain IMF_2 :

$$IMF_2 = \frac{1}{N} \sum_{i=1}^n E_1(r_1 + \varepsilon_1 E_1(n(i))) \quad (6)$$

The residual r_2 after removing the second-order IMF is:

$$r_2 = r_1 - IMF_2 \quad (7)$$

③ Repeat step ②, The $k+1$ th eigenmode component IMF_{k+1} and the k th residual component r_k can be obtained:

$$IMF_{k+1} = \frac{1}{N} \sum_{i=1}^N E_1(r_k + \varepsilon_k E_k(n(i))) \quad (8)$$

$$r_k = r_{k-1} - IMF_k \quad (9)$$

When the number of extrema in the IMF components is less than 2, K intrinsic mode functions IMF_k can be obtained. The original sequence is:

$$x = \sum_{k=1}^k IMF_k + R \quad (10)$$

In the formula, R represents the final residual.

2) The Principle of the Northern Goshawk Optimization Hyperparameter Algorithm

NGO, also known as the Northern Goshawk Algorithm, is a heuristic optimization algorithm. It is inspired by the hunting behavior of the northern goshawk. When hunting, the northern goshawk demonstrates high strategy and skill, capturing prey through various methods such as high-altitude swooping and covert approaching. These unique hunting behaviors have provided inspiration for the design of the Northern Goshawk Algorithm.

The Northern Goshawk Algorithm primarily simulates three fundamental behaviors of the goshawk: searching for prey, tracking prey, and attacking prey, in order to solve optimization problems. Here are the basic principles and steps of the algorithm:

- ① Initialization of the Population At the start of the algorithm, a group of "Northern Harriers" (a set of candidate solutions) is randomly generated, where each harrier represents a possible solution within the problem space.
- ② Searching for Prey In this stage, the harriers search for prey through random flights, that is, looking for possible optimized solutions. This process simulates the behavior of harriers observing the ground while flying in the air. The algorithm explores the solution space based on a certain strategy of search.
- ③ Tracking the Prey Once potential prey is discovered (i.e., a better solution is found), the harriers begin to track the prey. In this phase, the algorithm evolves towards a better solution using the currently known best solution, which can be achieved by simulating the harriers' circling above the prey, continuously reducing the distance to the prey.
- ④ Attacking the Prey After determining the location of the prey, the harriers attack through diving and other means. In the algorithm, this corresponds to conducting a more detailed search near the current best solution in the solution space, in order to find the global optimum or an approximate optimum solution.
- ⑤ Updating Position and Speed Based on the position of the prey (the position of the best solution) and the positions of other harriers, update the position and speed of each harrier, simulating the dynamic adjustment of the harrier population's flight strategy to adapt to changes in the prey.
- ⑥ Repeated Iteration Continue to perform the above operations until the termination criteria are met (i.e., reaching the maximum number of iterations or the quality of the solution is satisfactory). The Northern Harrier Algorithm utilizes the simulation of the hunting behavior of harriers to improve problem solutions, demonstrating excellent

global search capabilities and rapid convergence.

3) *The Principle of the CNN-GRU*

CNN-GRU is a hybrid neural network architecture that integrates the functionality of Convolutional Neural Networks (CNN) and Gated Recurrent Units (GRU).

CNN is a type of feedforward neural network characterized by its ability to automatically and effectively learn spatial hierarchical features from data. CNN first processes input data and extracts features through convolutional layers, then reduces the spatial dimensions of these features through pooling layers. This approach effectively captures local features, thereby reducing computational complexity and preventing overfitting.

As an improved version of the Recurrent Neural Network (RNN), GRU addresses the common issues of gradient decay or overshooting found in traditional RNNs through its unique "Gating Mechanism". By utilizing two separate yet interconnected key components—the "update gate" and the "reset gate", GRU can effectively regulate the flow of information and enables the model to capture not only short-term sequence patterns but also long-term information interactions.

In the CNN-GRU architecture, initial data feature extraction is completed through the convolutional neural network layer. These extracted high-level features contain the spatial information of the input data. Then, these features are fed into the GRU layer, which handles the temporal changes of these features, learning the time-series dependencies within the data. Finally, depending on the specific task requirements, fully connected layers output the final prediction results.

Through this combination, the CNN-GRU model can effectively process complex data with time-series features, achieving more accurate predictions and classifications.

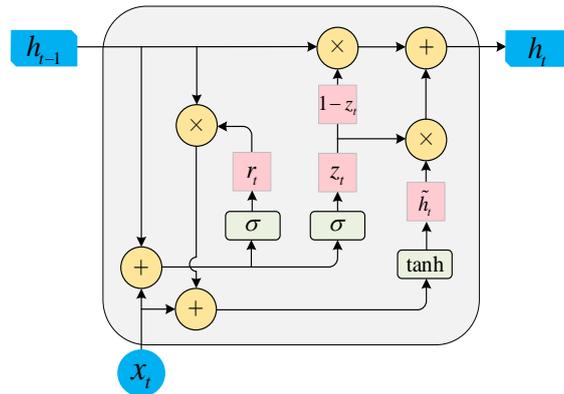


Figure 7 GRU Unit Structure

The main components of the GRU network include the reset gate and the update gate, as shown in Figure 7. In this model, the input x_t at each moment corresponds to the current hidden layer h_t output. The Sigmoid and tanh activation functions are used to process the input data respectively.

r_t , as a reset gate, can transform the current input x_t and the previous output h_{t-1} , as shown in formula (11). Adjusting the previous output h_{t-1} cannot control the previous input x_t , which determines the unit state requirement at this time point. The new input \tilde{h}_t is obtained through formula (12).

$$r_t = \sigma(W_r h_{t-1} + U_r x_t + b_r) \quad (11)$$

$$\tilde{h}_t = \tanh(W_h x_t + U_h r_t h_{t-1} + b_h) \quad (12)$$

The update gate z_t 's function is to control the influence of the previous time point input h_{t-1} and the new input

\tilde{h}_t , thereby adjusting the extent of its impact on the hidden layer output h_t . Z_t is obtained from equation (13). Z_t is used to control the new input \tilde{h}_t , while $1 - Z_t$ is used to manage the input h_{t-1} from the previous moment, obtaining the current moment output h_t through equation (14).

$$z_t = \tanh(W_z h_{t-1} + U_z x_t + b_z) \quad (13)$$

$$h_t = (1 - z_t) h_{t-1} + z_t \tilde{h}_t \quad (14)$$

In the formula, W 、 U are weight matrices, b is the bias.

The GRU network possesses a robust memory capacity; its predictions are contingent upon the comprehensive information patterns present throughout the temporal extent of the sample data, maintaining a significant correlation with future actual outcomes. The CNN-GRU neural network, an enhanced variant of the GRU, adeptly harnesses the convolutional neural network to discern unidentified features. It employs these extracted features to initialize hidden states and incorporates classification attributes into the input parameters. This refinement empowers the model to profoundly comprehend the sequential characteristics of the data, capturing both preceding and subsequent dynamics within the time series.

4) Self-attention mechanism

Self-attention, also known as intra-attention, is an indispensable attention mechanism employed within sequence models. It enables the capturing of dependencies among elements within a sequence, irrespective of the distance separating these elements. This mechanism has been extensively incorporated into the Transformer architecture, particularly within the domain of natural language processing. The fundamental steps comprising Self-attention are as follows:

Attention Scores: The self-attention mechanism operates by quantifying the extent to which each element should attend to other positions within the sequence. This is achieved through the computation of the similarity between a Query and all Keys, utilizing the dot product method, followed by a selective scaling process.

Softmax Layer: The attention scores, once derived, are subsequently processed through a Softmax layer. This treatment ensures that the aggregated attention score for each element's relation to all other elements within the sequence equates to 1. The application of the Softmax function transforms the elements into a probability distribution, providing a clear indication of the degree of attention each element affords to others within the sequence.

Weighted Value Vectors: The Value vectors corresponding to each element are then weighted and aggregated in accordance with the probability distribution produced by the Softmax layer. This synthesis generates the final output vector, which encapsulates the comprehensive contextual information of the entire sequence.

The refined exposition above maintains a rigorous, formal, and highly professional tone, adhering to standard language conventions, thus ensuring the content is both precise and authoritative.

For each element (i) in the sequence, calculate the dot product of Query(Q_i) with all Key(K_j), then obtain the attention weights w_{ij} through the Softmax function:

$$w_{ij} = \text{softmax}(Q_i \cdot K_j^T) = \frac{\exp(Q_i \cdot K_j^T)}{\sum_{k=1}^n \exp(Q_i \cdot K_k^T)} \quad (15)$$

Among them, n is the length of the sequence. Subsequently, these weights are used to calculate the sum of the weighted Value vectors, obtaining the output vector O_i for element:

$$O_i = \sum_{j=1}^n w_{ij} \cdot V_j \tag{16}$$

The advantageous nature of self-attention is rooted in its inherent flexibility and robust representational prowess, which substantially augments the model's efficacy in processing sequential data. This is particularly pronounced when grappling with dependencies over extended distances within the sequence..

II. ESTABLISH THE NGO-CEEDMAN-CNN-GRU-SE MODEL

The ensemble model is architected upon the Northern Goshawk Optimized CEEMDAN-CNN-GRU-SE framework (herein referred to as 'the model'), which amplifies its capacity for processing intricate time series data. Incorporating the self-attention mechanism, the model is endowed with the capability to attentively focus on the critical segments of sequence data during its analysis, thereby elevating both the overall efficacy and precision. The structure of this sophisticated model is elucidated below:

To mitigate forecasting discrepancies induced by fluctuations in microseismic b-value data, and to augment the predictive veracity of the model, an integrative approach encompassing multiple algorithms was meticulously executed, culminating in the CEEMDAN-CNN-GRU-SE ensemble model. The CEEMDAN methodology facilitates the decomposition of the time series data pertaining to microseismic b-values into a multitude of IMF (Intrinsic Mode Function) frequency components. These deconstructed components are subsequently channeled into the CNN-GRU-SE model in a sequential manner, with the paramount model hyperparameters being fine-tuned through the Northern Buzzard optimization algorithm to maximize the predictive outcomes. In the final stage, the projected IMF frequency components are superimposed, and the cumulative prediction results are synthesized to yield the ultimate b-value forecast.

The specific procedural schema is illustrated in Figure 8, and the sequential steps are as follows:

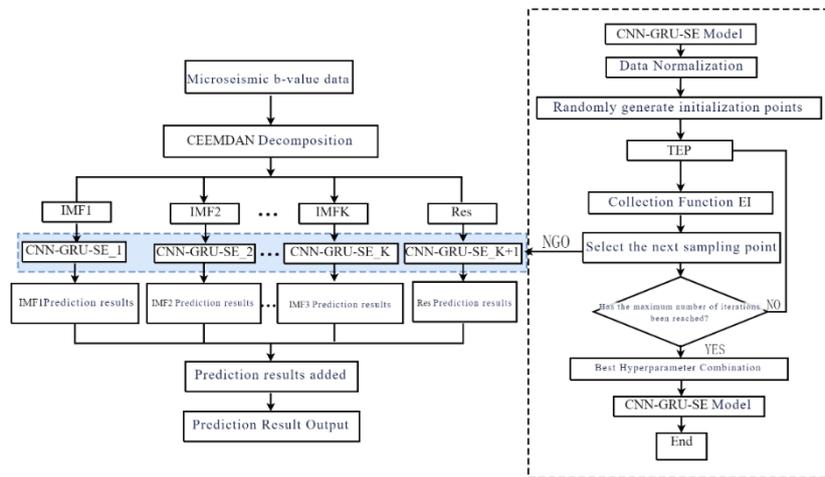


Figure 8 NGO-CEEDMAN-CNN-GRU combined model structure.

In the data preprocessing phase, the CEEMDAN method is utilized to analyze the time series data of the microseismic b-value, in order to address its randomness and instability. Through the deconstruction of the microseismic b-value sequence data, a succession of stable and methodical components can be extracted, known as IMFs, along with residual fractions.

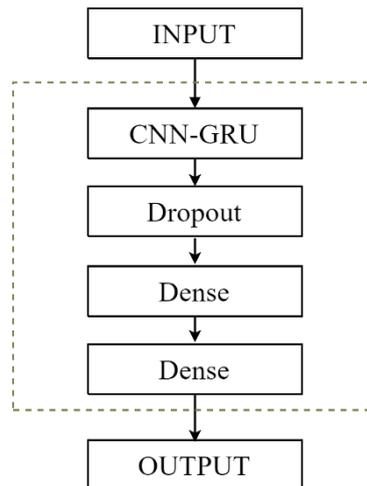


Figure 9 CNN-GRU Model Structure

(2) A CNN-GRU model was devised, and the CNN-GRU network was meticulously constructed within the MATLAB R2023a computational environment, as depicted in Figure 9. The architecture comprises a CNN-GRU layer, a Dropout layer, and two Dense layers. The Dropout layer serves as a safeguard against model overfitting, thus enhancing the capacity for generalization; the Dense layer, which is fully connected, processes the output dimension to yield the definitive predictive outcome. The activation function selected for the hidden layer of this model is relu, while the Adam optimizer is utilized, and the mean squared error (MSE) is adopted as the loss function.

(3) Normalize different components, and then adjust each part with a specific stride using a time window. As shown in Figure 10, the data $(x_{t-n}, x_{t-n-1}, \dots, x_{t-1})$ of the first n moments is used as input, and the data $n+1$ of the x_t moment is used as output. The CNN-GRU model is utilized and trained and tested.

$$\left\{ \begin{array}{cccc} x_1 & x_2 & \cdots & x_n \\ x_2 & x_3 & \cdots & x_{n+1} \\ \cdots & \cdots & \cdots & \cdots \\ x_{t-n} & x_{t-n+1} & \cdots & x_{t-1} \end{array} \right\} \left\{ \begin{array}{c} x_{n+1} \\ x_{n+2} \\ \cdots \\ x_t \end{array} \right\}$$

Figure 10 Slide Window Data Division

(4) To enhance the learning efficacy of the model, it is essential to accurately calibrate the optimal count of sampling points and the iteration frequency within the Northern Goshawk Optimization Algorithm. The number of neurons in the hidden layer, the frequency of iterations, and the step size of the sliding window are all tuned as hyperparameters targeted for optimization, alongside establishing the most suitable ranges for these hyperparameters. By refining the aforementioned parameters, the model training process is expedited, enabling the model to converge upon the optimal solution with increased alacrity.

The CNN-GRU-SE model established in the previous stage is used to train various feature training data and their initial setting samples, with the average square root error in the test set being used as the result of the objective function. By continuously adjusting hyperparameters, the output value for each feature is obtained. The best combination of hyperparameters is formed based on the learning rate of the CNN-GRU-SE model, the number of

hidden layer neurons, the number of iteration rounds, and the stride of the slide.

(5) The learning and prediction tasks are carried out through the CNN-GRU-SE model and its corresponding modules. The frequency components of b-values obtained from each prediction are accumulated. Subsequently, the root mean square error (RMSE) and mean absolute percentage error (MAPE) are used to evaluate the prediction accuracy of the forecasting model. The formulas for each evaluation metric are as follows, where x_i represents the true value and \hat{x}_i represents the predicted value:

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - \hat{x}_i| \quad (17)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{n}} \quad (18)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - \hat{x}_i}{x_i} \right| \times 100\% \quad (19)$$

Through such a design, the model not only leverageThrough such a design, the model not only leverages the advantages of CEEMDAN, CNN, GRU, but also enhances them with the self-attention mechanism, making the model more precise and efficient in analyzing and predicting complex time series data.

III. MODEL EFFECTIVENESS ASSESSMENT

From March 2, 2019, to January 10, 2020, the b-value of microseismic events in the area near Zhongyang Alley was used as a dataset for a total of 315 days. The first 80% of the microseismic b-value data was used as the training set, with the remaining 20% serving as the test set for constructing and evaluating the model. As previous research has indicated, the data for the b-value exhibits significant instability and volatility. Running the model with the undecomposed original sequence can lead to a decrease in the accuracy of the model. Therefore, the model adopted CEEMDAN to denoise and decompose the b-value data, ultimately breaking it down into several IMF frequency components and one Res residual component.

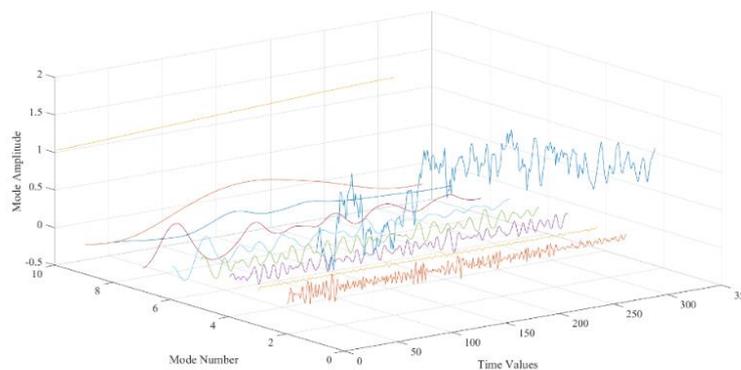


Figure 11 Time Series Decomposition Chart for Value

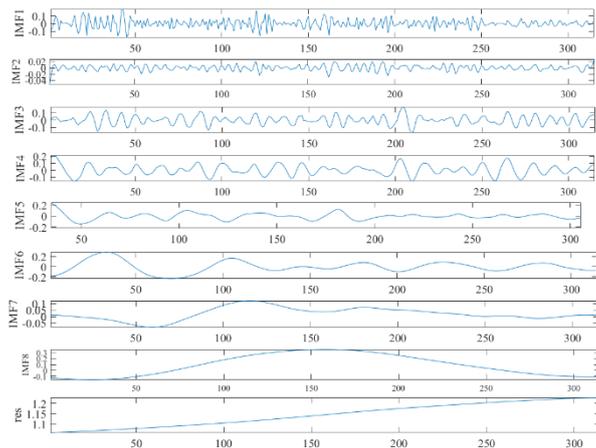


Figure 12 b-value IMF Component Chart

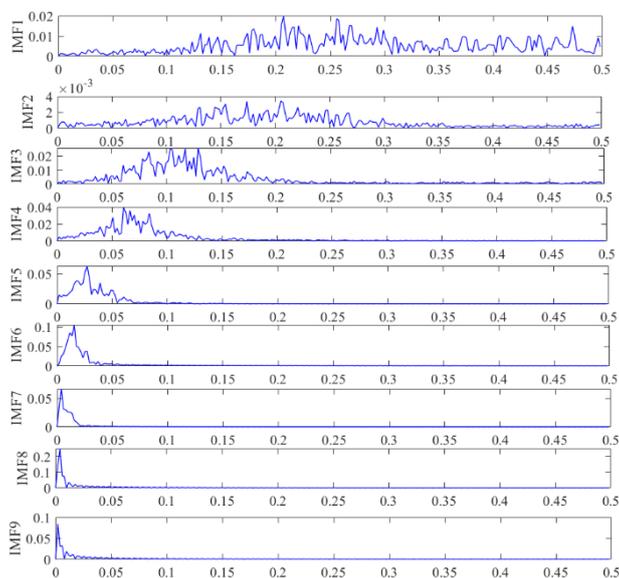


Figure 13 IMF Component Spectrum Diagram

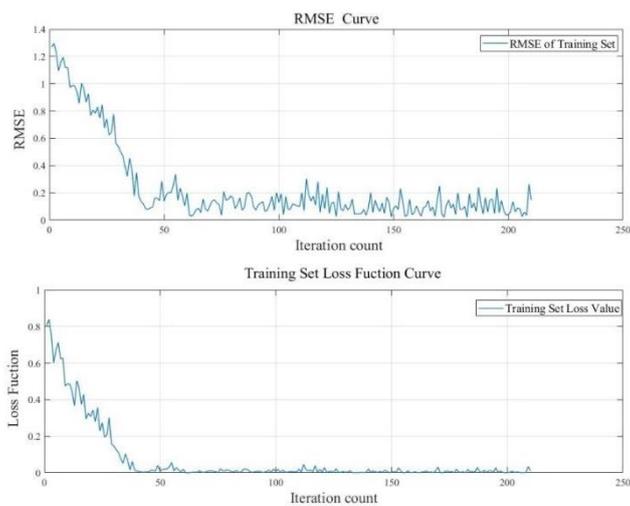


Figure 14 Training Set Root Mean Square Error Curve

With the execution of the steps outlined above, each augmented stability and each decomposed component demonstrates augmented stability and diminished variability, thereby fostering an environment conducive to the effective training of neural network models.

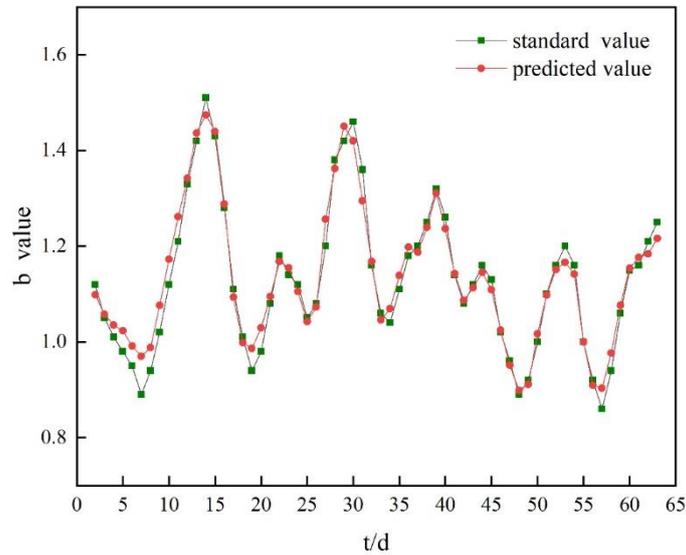


Figure 15 Comparison between Predicted and Actual Values

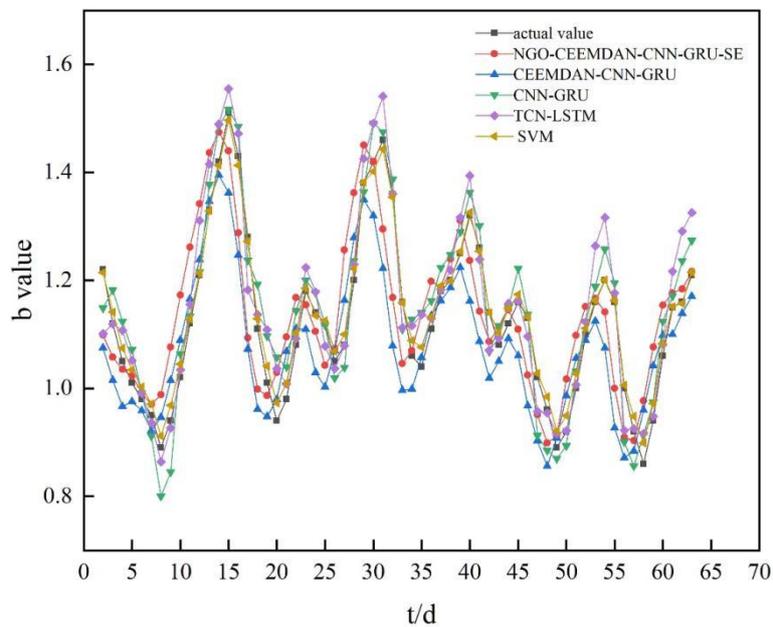


Figure 16 Comparison of Predicted Values from Different Models

Figures 15 and 16 clearly demonstrate that the optimized performance of the NGO-CEEMDAN-CNN-GRU-SE model more accurately portrays the fluctuations of the b-value. The CEEMDAN method exhibits high precision in one-step ahead forecasting, with no discernible lag, and a relatively stable prediction error, free from significant fluctuations. Notably, the model maintains an excellent tracking capability even amid substantial data shifts. In contrast, other models such as CNN-GRU, TCN-LSTM, SVM, etc., do not fare as well; they suffer from higher error rates and pronounced time lag issues, leading to inferior prediction results. The superior performance of the

CEEMDAN method can be attributed to its frequency decomposition process, which uncovers the periodic patterns of the original data across various temporal frequency bands. This approach mitigates the inherent instability and stochastic fluctuations of the microseismic b-value, eliminates prediction lags in the CNN-GRU model, and enhances forecast accuracy. Consequently, this refined model is capable of predicting future b-values and providing precise early warning signals for potential rock burst risks.

For the prediction data of different models, we use For the prediction data of different models, using three evaluation methods: MAE, RMSE, and MAPE to assess their accuracy. The relevant data is detailed in Table 2. As can be seen from the table, compared to CEEMDAN-CNN-GRU, the model in this study has reduced errors in predicting future b values, indicating that the optimization strategy has played a certain role in improving the prediction accuracy of the model. In summary, the model proposed in this study can effectively predict the changes in b value over a period of time in the future. The values of MAE, RMSE, and MAPE are all the lowest. This shows that the model can effectively learn the deep feature information of microseismic b value at historical moments for prediction, thereby improving the accuracy of prediction and providing an important basis for the early warning of rock bursts.

Table 2 Comparison of Several Model Evaluation Indicators

Model	R ²	MAE	RMSE	MAPE
NGO-CEEMDAN - CNN-GRU-SE	0.9638	0.022	0.028	2.09%
CEEMDAN - CNN-GRU	0.8356	0.052	0.061	4.49%
CNN-GRU	0.6274	0.073	0.092	6.59%
TCN-LSTM	0.6792	0.069	0.085	6.22%
SVM	0.7182	0.069	0.080	6.24%

The comparative analysis of different models highlights the superior performance of this study's model in terms of accuracy and reliability, indicating that it can be further refined and enhanced to better suit more intricate geological conditions and mining environments.

IV. CONCLUSION

Research has been conducted to establish an early warning model for rockburst based on deep recurrent neural networks, which predicts the occurrence of rockburst events by analyzing the b-value in microseismic monitoring data. The development and testing results of this model indicate that using the b-value of microseismic events as a predictive indicator can effectively forecast the potential risk of rockbursts, holding significant practical importance for ensuring mine safety production.

1. The energy of microseismic events is closely associated with the occurrence of rockbursts: An extensive analysis of microseismic event data has revealed a strong correlation between the b-value and the incidence of rockburst events. Calculated from the G-R relationship, the b-value is monitored, and it has been observed that before large-energy microseismic events occur, the corresponding b-value and its surrounding time periods' b-values are all in a diminished state, falling below the average b-value baseline. This average b-value line serves as an early warning threshold for significant events, with fluctuations in the b-value trend acting as a predictive indicator for potential rockbursts.

2. The early warning model grounded in deep recurrent neural networks exhibits high efficacy: Substantiated by empirical data testing, the model has demonstrated remarkable prediction accuracy and dependability, adeptly forecasting the emergence of high-energy microseismic events, thus providing robust support for preemptive warnings in mine safety.

3. The model presented in this paper achieves the lowest MAE (Mean Absolute Error), RMSE (Root Mean Square

Error), and MAPE (Mean Absolute Percentage Error) for future 1 to 3 step predictions of the b-value, showcasing commendable predictive performance. It is clear that the model possesses the capability to effectively learn and capture the historical patterns of microseismic b-value changes, offering high precision in predictions, and supplying a reliable foundation for rockburst early warning systems.

4. The model offers practical application value: By facilitating timely and preventative measures, this proactive warning model provides essential technical support for safe mining operations, thereby mitigating the risks associated with rockburst events.

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.

ACKNOWLEDGMENTS

The authors are grateful to Prof. Enke HOU for useful information provided.

REFERENCES

- [1] Ministry of Natural Resources, PRC. “*China Mineral Resources 2023*”. Geology Press, Beijing, 2024-04-02.
- [2] D Linming, T Xinyuan, C Anye, et al. Present situation and problems of coal mine rock burst prevention and control in China. *Journal of China Coal Society*, 2022, 47(1):152-171. DOI:10.13225/j.cnki.jccs.YG21.18
- [3] Q Qingxin, L Yizhe, Z Shankun, et al. Seventy years development of coal mine rockburst in China: establishment and consideration of theory and technology system. *Coal Science and Technology*, 2019, 47(9): 1-40. DOI:10.13199/j.cnki.cst.2019.09.001.
- [4] L Hang, C Dongke, M Debing. Current status of deep mining and disaster prevention in China. *Coal Science and Technology*, 2016, 44(1):39-46. DOI:10.13199/j.cnki.cst.2016.01.007.
- [5] H Wei, Q Xuan-ye, C Yi-tao, et al. Research of Risk of Rock Burst Based on Micro-seismic Monitoring and Multi-factor Coupling. *Coal Technology*, 2016, 35(1):110-112. DOI:10.13301/j.cnki.ct.2016.01.043.
- [6] C Jie, G Jingkuan, P Yuanyuan, et al. Machine learning method for predicting and warning of rockbursts. *Journal of Mining and Strata Control Engineering*, 2021, 3(1):57-68. DOI:10.13532/j.jmsce.cn10-1638/td.20200922.001.
- [7] H Shengquan, H Xueqiu, S Dazhao, et al. Multi-parameter integrated early warning model and an intelligent identification cloud platform of rockburst. *Journal of China University of Mining & Technology*, 2022, 51(5):850-862. DOI:10.13247/j.cnki.jcumt.001440.
- [8] L Haitao, Z Haikuan, Q Qingxin. A CNN-GRU rock burst risk analysis model considering micro-seismic precursor characteristics. *Safety in Coal Mines*, 2023, 54(7):41-49. DOI:10.13347/j.cnki.mkaq.2023.07.006.
- [9] Y Xu, L Yapeng, C Anye, et al. A spatio-temporal prediction method for coal burst based on the fusion of microseismic multidimensional information. *Journal of Mining & Safety Engineering*, 2024, 1-11. DOI : 10.13545/j.cnki.jmse.2023.0453.
- [10] S Xuebo, L Ning, W Yuanjie, C Fabing, L Yan. Application of time series ARIMA model in coal mine ground sound monitoring system. *Coal Engineering*, 2023, 55(10):111-117. DOI:10.11799/ce202310019.
- [11] Q Changkun, Z Wusheng, J Haibin et al. A Method for Predicting the Time Series of Microseismic Events in Coal Mines Based on Modal Decomposition and Deep Learning. *Journal of China Coal Society*, 2024, 1-16.
- [12] L Yaoqi, C Anye, W Changbin, et al. Prediction method of coal burst risk based on focal mechanism and location error calibration. *Journal of China Coal Society*, 2023, 48(5):2065-2077. DOI:10.13225/j.cnki.jccs.2023.0093.
- [13] L Yanhua, S Fenbo, X Zidian, W Junfeng. Research on prediction method of rock burst based on LSTM model. *China Mining Magazine*, 2023, 32(5):88-95. DOI:10.12075/j.issn.1004-4051.2023.05.017.
- [14] C Jie, D Junsheng, P Yuanyuan, J Deyi, Q Qingxin. “Dual-driven” intelligent pre-warning framework of the coalburst disaster in coal mine and its engineering application. *Journal of China Coal Society*, 2022, 47(2):791-806. DOI:10.13225/j.cnki.jccs.XR21.1897.

- [15] H Yongliang,W Suping,F Yuping,C Xuefang,S Dali.Early-warning and soft structure prevention technology of rock burst risk based on multi-source information fusion.Safety in Coal Mines,2023,54(7):78-84. DOI:10.13347/j.cnki.mkaq.2023.07.010.
- [16] L Baolin;W Enyuan;H Shaobin;M Ali.Minerals.Evaluation of Vibration Effect Caused by Carbon Dioxide Phase-Transition Fracturing Based on the Hilbert–Huang Transform. 2022. DOI:10.3390/MIN12020242.
- [17] J Baoxin,C Hao,P Yishan, C Yang.Rock Burst Prediction Technology of Multi-parameters Synthetic Index.Journal of Disaster Prevention and Mitigation Engineering, 2019, 39 (02): 330-337. DOI:10.13409/j.cnki.dpme.2019.02.019.
- [18] P Yishan,S Yimin,Z Chenli, R He,X Hailiang.Localization method of coal rock deformation for rock burst prediction.Journal of China Coal Society,2023,48(1):185-198. DOI:10.13225/j.cnki.jccs.2022.1375
- [19] L Junping,G Tingting,F Jiayu, W Haiquan.Research progress on prevention and control of mine earthquake and rockburst.China Safety Science Journal,2024,34(1):85-93.DOI:10.16265/j.cnki.issn1003-3033.2024.01.0103.
- [20] P Junfeng,X Yongxue,W Shuwen,et al. Technical difficulties and emerging development directions of deep rock burst prevention in China. Journal of China Coal Society,2024,49(3):1291–1302. DOI:10.13225/j.cnki.jccs.2023.1480.