



OPEN ACCESS

EDITED BY

Xin Yin,
City University of Hong Kong, Hong
Kong SAR, China

REVIEWED BY

Huan Sun,
Hainan University, China
Yunpeng Zhang,
China University of Geosciences
Wuhan, China

*CORRESPONDENCE

Haikuan Zhang,
✉ 820818414@qq.com

RECEIVED 27 March 2025

ACCEPTED 28 April 2025

PUBLISHED 29 May 2025

CITATION

Zhang X, Li G, Chen Y, Wang H, Zhang H, Li H,
Du W, Li X, Xu X and He Y (2025) A
prototype-based rockburst types and risk
prediction algorithm considering intra-class
variance and inter-class distance of
microseismic data.
Front. Earth Sci. 13:1601090.
doi: 10.3389/feart.2025.1601090

COPYRIGHT

© 2025 Zhang, Li, Chen, Wang, Zhang, Li, Du,
Li, Xu and He. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

A prototype-based rockburst types and risk prediction algorithm considering intra-class variance and inter-class distance of microseismic data

Xiufeng Zhang¹, Guoying Li¹, Yang Chen¹, Hao Wang¹,
Haikuan Zhang^{2*}, Haitao Li², Weisheng Du², Xiao Li²,
Xuewei Xu² and Yuze He²

¹Coal Industry Management Department, Shandong Energy Group Co., Ltd., Jinan, Shandong, China,

²Deep Mining and Rock Burst Research Branch, Chinese Institute of Coal Science, Beijing, China

The prediction and classification of rockburst risk based on microseismic data is the premise of preventing rockbursts during deep mine excavation. By reviewing previous studies, this paper finds two problems that hinder the rockburst prediction: 1) there is a lack of research on the distribution features of monitoring data on the main controlling factors of rockbursts; 2) there is no research on the intra-class variance and inter-class gap of microseismic data. Based on the typical rockburst risk events, a quantitative information model of geology and mining is constructed. The relationship between the spatial-temporal distribution characteristics of microseismic data before a rockburst and the main controlling factors of a rockburst is studied. The results show that the distribution features may be different for the same type of microseismic (MS) and rockburst events, and different types of events may show similar distribution features. Therefore, based on the quantitative study of the relationship between the performance of a deep learning prediction algorithm and a rockburst prediction vector, a rockburst risk and type prediction algorithm based on a convolutional neural network (CNN)-gated recurrent unit (GRU) model with prototype-based prediction is proposed. The CNN-GRU model can produce prediction vectors by fusing implicit and explicit information extracted from the original MS data and early warning indicators. Cross-entropy loss, vector-prototype contrastive loss, and vector-prototype contrastive loss are proposed to automatically control the intra-class variance and inter-class gap of prediction vectors belonging to different rockburst risks and types. Many experiments show that the performance of the proposed CNN-GRU model with prototype-based prediction is superior to other algorithms in the prediction of rockburst risks and types based on MS data.

KEYWORDS

rockburst prediction, rockburst types, deep learning, microseismic data, prototype learning

1 Introduction

As underground mining and excavations continue to expand at a rapid pace, a significant engineering challenge arises in the instability of the surrounding rock masses (Aydan et al., 2017). Rockburst, a representative instability phenomenon, is caused by the abrupt release of accumulated elastic strain energy, posing a grave threat to the safety of workers and causing extensive damage to underground engineering structures (Basnet et al., 2023). A significant number of researchers are committed to addressing this challenge by elucidating the occurrence mechanism (Askaripour et al., 2022; He et al., 2023), enhancing prediction accuracy (Basnet et al., 2023; Pu et al., 2019), and developing effective control measures (Li et al., 2019; He et al., 2018). Rockburst prediction aims to generate accurate signals prior to the occurrence of these disasters, serving as a prerequisite for controlling and managing the rockburst hazards. Nevertheless, due to its unpredictable emergence and numerous influencing factors, rockburst prediction remains a challenging task that has yet to be fully resolved.

Rockburst prediction is conventionally classified into long-term prediction and short-term prediction (Liang et al., 2020). Long-term rockburst prediction endeavors to utilize rock mechanical parameters for constructing a prediction model with the aim of assessing the rockburst probability and types of diverse surrounding rock masses under assorted field conditions. This task is customarily accomplished during the initial stage of engineering design or excavation (Liang and Zhao, 2022). The objective of short-term rockburst prediction, on the other hand, is to predict the time, types, and damage magnitude of dangerous rockburst events by conducting dynamic and static analyses on real-time monitoring data. It is typically carried out during the excavation period (Jinqiang et al., 2021). To attain satisfactory outcomes, researchers are dedicated to applying diverse methods for rockburst prediction, such as empirical analytical (He et al., 2018; Yang et al., 2018), experimental (Hu et al., 2023; Cheng et al., 2023), numerical (Wang et al., 2021; Manouchehrian and Cai, 2018), intelligent (Adoko and Zvarivadza, 2018; Xue et al., 2023), and expert system (Li et al., 2020) methods. Although each rockburst prediction method has its own advantages, in contrast to intelligent methods, traditional prediction methods hinge on extensive expert experience and meticulous judgment. Therefore, the machine learning method represents a promising alternative and has been adopted by numerous researchers to dissect and handle the intricate and nonlinear process of rockburst prediction.

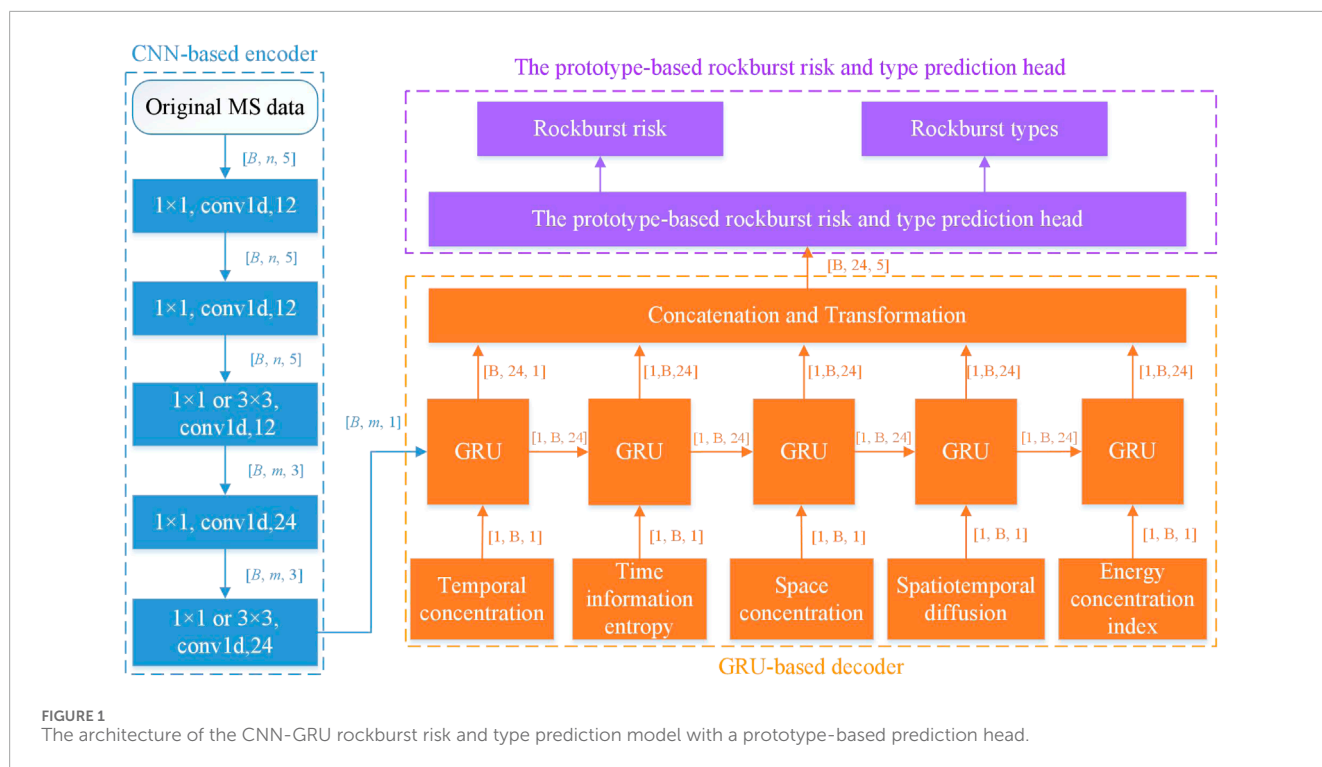
Microseismic (MS) monitoring is a widely acknowledged and highly effective tool for identifying the dangerous signals and key controlling factors of rockburst types for short-term rockburst prediction. It is capable of monitoring the occurrence of rockbursts by extracting valuable signals that propagate from the fracturing process of rock masses. Recently, by capitalizing on the capabilities of machine learning in handling nonlinear problems, scholars have directed their attention to rockburst prediction based on MS data through the application of existing machine learning algorithms. Such algorithms encompass support vector machine (SVM) (Ji et al., 2020; Jin et al., 2022), convolutional neural network (CNN) (Dong et al., 2023; Zhang et al., 2021; Yin et al., 2021a),

variants of RNN (Hu et al., 2023; Di et al., 2023a; Di et al., 2023b), convolutional long short-term memory (ConvLSTM) (Chen et al., 2023; Ma et al., 2021), and ensemble-learning (Liang et al., 2020; Yin et al., 2021b; Liang et al., 2021), among others (Yin et al., 2024a; Yin et al., 2024b; Cheng et al., 2024; Yin et al., 2021c). Zhang et al. (2021) and Yin et al. (2021a) have conducted an exploration of the key microseismic indexes that can characterize the development process of rockbursts and have applied the refined convolutional neural network (CNN) to predict rockbursts. In order to depict the spatiotemporal relationship within microseismic data and process the spatiotemporal indexes for rockburst prediction, Chen et al. (2023) developed a deep learning model founded on a ConvLSTM to forecast short-term rockburst risks. Concurrently, the ensemble-learning methodologies (Liang et al., 2020; Yin et al., 2021b; Liang et al., 2021) have also been employed to acquire a highly potent rockburst prediction model relying on MS parameters.

Previous research on the rockburst prediction first analyzed the distribution features of microseismic data before a rockburst occurred. Then, the rockburst prediction indexes are presented, and the machine learning algorithm is used to describe the relationship between the rockburst prediction indexes and the future rockburst risks and types. However, these studies ignore the influence of the main controlling factors of rockbursts on the distribution of MS events before a rockburst occurs. In the rockburst prediction and classification process, there is no quantitative analysis of the influence of the intra-class variance and inter-class gap of prediction vectors on the prediction accuracy. The prediction vectors are the result of rockburst prediction indexes or MS data processing by deep learning encoders.

A quantitative information model of geology and mining is constructed based on the different types of rockburst risk events. The temporal and spatial distribution characteristics of MS data before a rockburst, in the main controlling factors of rockburst (i.e., geological structure, roof, coal pillar, and high-stress coal mass), are studied. The results show that the distribution features of microseismic and rockburst events of the same type may differ, and different types of events may show similarities in space-time distribution. The relationship between the performance of a deep learning prediction algorithm and the prediction vector of rockbursts is studied. The rockburst prediction model must accurately distinguish the differences between different types of data and accurately describe the intra-class variance of monitoring data.

A rockburst risk and type prediction method based on a convolutional neural network (CNN)-gated recurrent unit (GRU) and prototype learning prediction head is proposed. The CNN-GRU model can generate prediction vectors by integrating both implicit and explicit information derived from original microseismic (MS) data and prediction indexes. Through the use of cross-entropy loss, vector-prototype contrastive loss, and inter-vector-prototype contrastive loss, the model can autonomously manage the intra-class variance and inter-class distance of prediction vectors corresponding to different rockburst risks and types. Extensive experimental results demonstrate that our proposed CNN-GRU model with prototype-based prediction outperforms alternative algorithms in forecasting rockburst risks and types using MS data.



2 Spatial–temporal and intensity evolution of rockbursts in MS data

Although the factors influencing rockbursts are complex, four controlling factors affecting the structure and stress of rockbursts are commonly recognized: geological structure, hard rock strata, coal pillars, and mining depth. The geological and mining data related to these four factors from typical rockburst events are first quantified. For each working face, the mining progress over a 10-day sliding window throughout the entire mining process is calculated. The geological and mining information models corresponding to different working faces are constructed. Then, by projecting the MS data to the quantified geological and mining information model, MS events and rockbursts are simply divided into four types, that is, the hard rock roof type, geological structure type, coal pillar type, and mining depth type, as shown in Figure 1. Some events are caused by more than one controlling factor.

Figure 2 shows that the distribution of different types of MS events is disordered and irregular within 10 days before the rockburst. Even for the same type of events, microseismic events are clustered in several central regions and do not show similar distribution characteristics. Meanwhile, the same type of MS events is clustered in several central regions and do not show similar distribution characteristics. During long-term mining, most MS events show the characteristics of irregularity and relatively uniform distribution and occur in the area where the main control factors are active. When MS events occur intensively in some special areas, it indicates that there is a rockburst risk in the area. When MS events occur intensively in some special areas during the short term, this indicates that there is a rockburst risk in these areas.

The sequential expansion of the time and MS energy across the entire working face is illustrated in Figure 3. Over the long term, it is obvious that the main active controlling factors of a rockburst are the same as those of most MS events. This is because there are structural planes that hinder energy propagation among these four main controlling factors, and the occurrence area of most MS events can be used to determine the destination or main path of energy propagation. The variation of MS energy over time scales is disordered. From a short-term perspective within a rockburst occurrence, the energy and frequency of MS events have increased slightly, showing a relatively active state.

In summary, the MS events can be divided into four types according to the time and space projection of MS events on the controlling factors. For the entire mining process, all types of MS events are disordered and irregular in terms of temporal and spatial characteristics. When the same type of MS events evolves from disordered and scattered to ordered and intensive, the main controlling factor area may have a rockburst hazard. However, even in the short term of rockburst occurrence, MS events are clustered within several centers on the time and space scale. Therefore, it is difficult and complex to summarize the temporal and spatial characteristics of the same type of MS data and accurately distinguish between different types.

3 Relationship of prediction head and rockburst prediction vectors

In this paper, the content of rockburst prediction includes the risk and type. There are two rockburst risk prediction results: dangerous and non-dangerous. There are four rockburst type prediction results: gravity type, coal pillar type, roof type, and

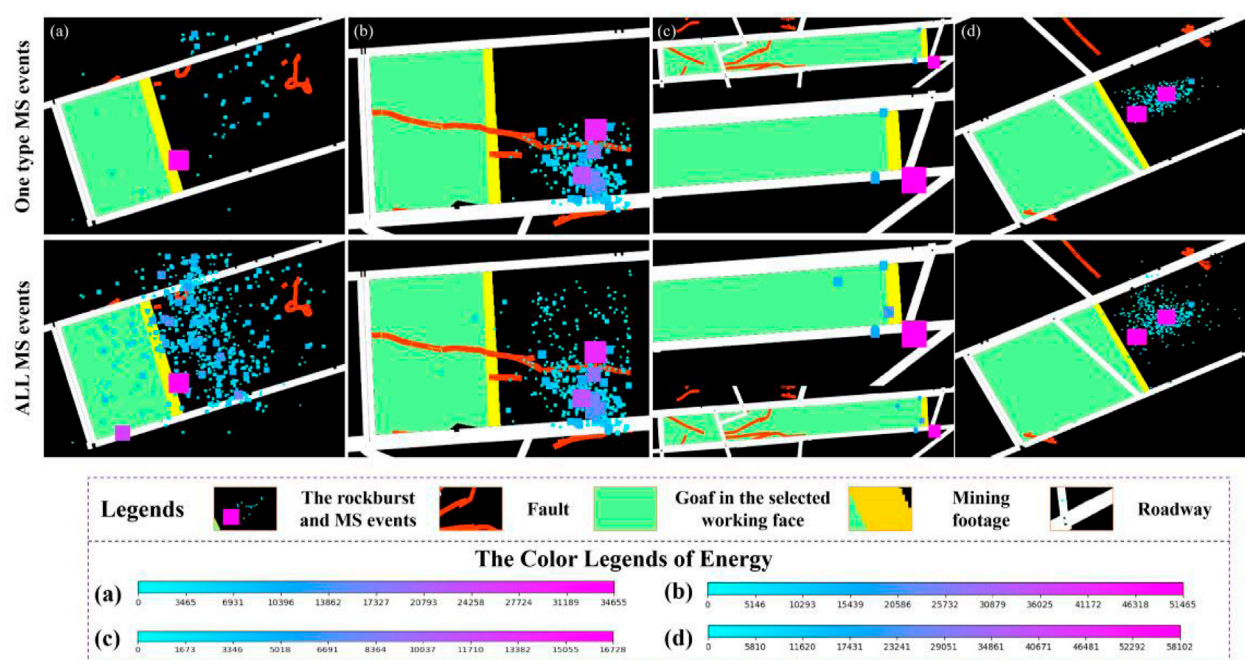


FIGURE 2

Rockbursts and MS events caused by different main controlling factors within the working face. The MS data of all figures are within 10 days of the rockburst occurrence. The first-row figures show the dangerous MS events influenced by the (a) hard rock roof, (b) geological structure, (c) coal pillar, or (d) gravity or mining depth factors. The second row figures show the projection results of all MS data.

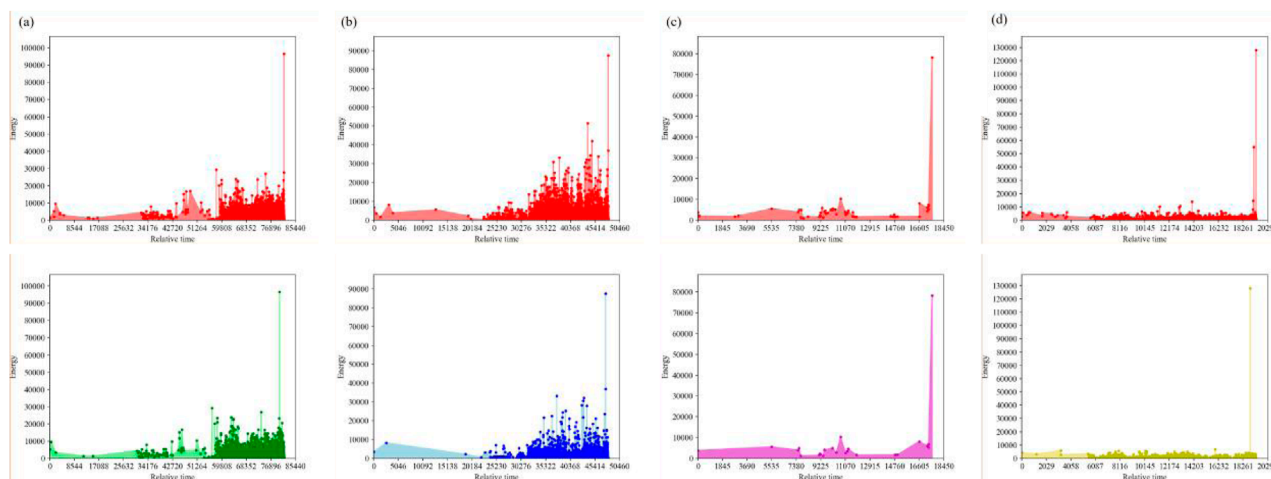


FIGURE 3

Time sequence exhibition of various MS events within the working faces where a (a) hard rock roof type rockburst, (b) geological structure type rockburst, (c) coal pillar type rockburst, and (d) mining depth or gravity type rockburst occurs. Selected working faces are the same as those in Figure 3. The MS event with the maximum energy value is the rockburst.

tectonic type. For the machine learning-based rockburst risk and type prediction method, the algorithms mainly consist of the encoder, decoder, and prediction head. The encoder is mainly responsible for analyzing the input MS data and extracting data features. For most previous rockburst prediction methods, the function of an encoder was replaced by the calculation formulas of prediction indexes. The decoder analyzes the feature

space of the data features and processes them into vectors with rockburst risk and type information. The final prediction head maps such vectors to the probabilities of rockburst risk and rockburst types.

This section takes rockburst risk prediction as an example to study the distribution requirement of decoder output vectors and representative vectors in the prediction head. For the traditional

rockburst risk prediction head, it is reasonable to assume that there are two representative vectors, that is, a non-dangerous representative vector w_1 and a dangerous representative vector w_2 . After the rockburst prediction vector v is obtained, the essence of this problem is to determine the probability that v belongs to risk class C according to the relationship between v and the representative vector w_c in the prediction head, which can be regarded as follows:

$$f(\omega_c^T v + b_c) = \text{const} \quad (1)$$

The least squares error between the dangerous probability calculated based on the vector v and the target y_i can be written as follows:

$$L = \frac{1}{2} \sum_{n=1}^N (\omega_c^T v_n + b_c - y_n)^2 \quad (2)$$

where N is the total number of samples. To simplify the calculation, the C label is changed to $y_c = N/N_1$, and the other label is changed to $y_{\bar{c}} = -N/N_2$. To obtain the maximum likelihood estimate for b_c , setting L with respect to b_c to 0, the following result can be obtained:

$$b_c = -\frac{1}{N} \omega_c^T (N_c m_c + N_{\bar{c}} m_{\bar{c}}) \quad (3)$$

where m_c and $m_{\bar{c}}$ are the two category centers of dangerous and non-dangerous vectors. Setting the derivatives of L with respect to w_c :

$$\frac{\partial L}{\partial \omega_c} = \left(\sum_{n \in C} v_n v_n^T + \sum_{m \in \bar{C}} v_m v_m^T - N_c m_c m_c^T - N_{\bar{c}} m_{\bar{c}} m_{\bar{c}}^T \right) \omega_c - N(m_c - m_{\bar{c}}) \quad (4)$$

where $m = (N_c m_c + N_{\bar{c}} m_{\bar{c}})/N$. In order to analyze the data distribution requirements of vector v data, the intra-class variance is

$$\sum_{n \in C} [\omega_c^T (x_n - m_c)]^2 + \sum_{n \in \bar{C}} [\omega_c^T (x_n - m_{\bar{c}})]^2 = \omega_c^T S_w \omega_c \quad (5)$$

where k is the class index. The S_w is

$$S_w = \left\{ \sum_{n \in C} (v_n - m_c)(v_n - m_c)^T + \sum_{n \in \bar{C}} (v_n - m_{\bar{c}})(v_n - m_{\bar{c}})^T \right\} \quad (6)$$

The inter-class distance can be expressed as

$$(m_{\bar{c}} - m_c)^2 = \omega_c^T S_B \omega_c \quad (7)$$

where the S_B is

$$S_B = (m_{\bar{c}} - m_c)(m_{\bar{c}} - m_c)^T \quad (8)$$

According to Equation 6, the following equation can be obtained.

$$\sum_{n \in C} v_n v_n^T + \sum_{m \in \bar{C}} v_m v_m^T = S_w + N_c m_c m_c^T + N_{\bar{c}} m_{\bar{c}} m_{\bar{c}}^T \quad (9)$$

Combining Equation 8 and Equation 9, Equation 4 can be transformed to

$$\sum_{n \in C} v_n v_n^T + \sum_{m \in \bar{C}} v_m v_m^T = S_w + N_c m_c m_c^T + N_{\bar{c}} m_{\bar{c}} m_{\bar{c}}^T \quad (10)$$

Setting Equation 10 to zero, the following equation can be obtained:

$$\omega_c \propto S_w^{-1}(m_{\bar{c}} - m_c) \quad (11)$$

Due to the differences in the monitoring data, S_w is positive and not proportional to the unit matrix. Therefore, in the task of rockburst risk prediction, the vector v should not only maximize the variance between categories ($m_{\bar{c}}$ and m_c) but also minimize the variance within each category S_w . These two objects are interactive. This conclusion is also applicable to the prediction of rockburst type. If there is no proper training method, reducing the intra-class variance of vectors within the same class will also lead to the reduction of inter-class distance of vectors. Therefore, in this paper, the decoder output vectors of rockburst risk and type are controlled using a prototype-based prediction head that can accurately control the intra- and inter-class variances of vectors.

4 CNN-GRU model with the prototype-based prediction head

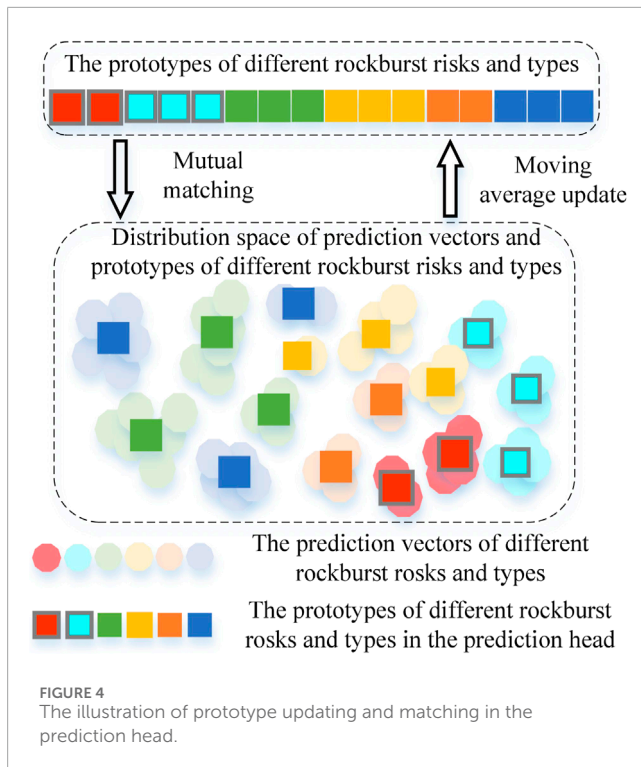
According to the analysis in Section 2 and Section 3, the distribution of rockburst data is very complicated, and it is difficult to describe the distribution of rockburst prediction vectors belonging to one rockburst risk class or type by a representative vector in the traditional classifier head. Moreover, the traditional prediction headers based on single or multi-layer perceptrons are updated by backpropagation. This results in the distribution space of the rockburst prediction vectors that cannot be reasonably controlled. Therefore, a novel CNN-GRU algorithm for comprehensively analyzing the original MS data and rockburst prediction indexes is presented to project the MS data to rockburst risk and type prediction vectors by automatically analyzing and fusing the spatiotemporal distribution features of MS data. Then, a prototype-based rockburst risk and type prediction head that can control the inter-class distance and intra-class variance of the prediction vectors is constructed.

4.1 The encoder-decoder based CNN-GRU algorithm

As mentioned in Section 2, the distribution characteristics of MS events in different stages of rockburst development are very complicated. Therefore, in previous studies, the MS data are first transformed into prediction indexes before inputting the machine learning model. In this paper, the CNN-GRU algorithm outputs the prediction vectors by comprehensively analyzing the original MS data and the prediction indexes, as shown in Figure 4.

The encoder consists of five one-dimensional (1D) CNN-based layers. Each 1D CNN layer contains a convolution layer, a batch normalization layer, and a rectified linear unit (ReLU) activation function. The number of convolution kernels in each layer is 12, 12, 12, 24, and 24, respectively. The input of the CNN-based encoder is the original MS data with a standardized time interval of two adjacent events, X, Y, and Z coordinates, and energy. If the number of analyzed MS data elements is n and the dimension of the prediction vector is 24, the input and output dimensions of the encoder are [batch size, n , 5] and [1, batch size, 24].

The decoder consists of five GRU-based modules. The hidden layer parameter of the first GRU module is the encoder output. The input of each GRU module is made up of different rockburst



prediction indexes. The decoder outputs are the prediction vectors whose dimension is $[1, \text{batch size}, 24]$. In this model, the rockburst prediction index consists of standardized temporal concentration Q_T , the time information entropy Q_t , space concentration Q_D , spatiotemporal diffusion d_s , and energy concentration index Q_E . The temporal concentration Q_T can be formalized as

$$Q_T = \text{Var}(T_n) / \Delta \bar{T}_n \quad (12)$$

where Q_T , $\text{Var}(T_n)$ and $\Delta \bar{T}_n$ are the temporal concentration, the variance, and the mean value of the time interval of the last n MS events.

Therefore, the time information entropy Q_t is used to describe the aggregation degree of MS events in the time series, reflecting the disorder or order in the evolution of MS time.

$$Q_t = \frac{-(1/n) \sum_{i=1}^n p_i \ln p_i}{\ln(n-1)} \quad (13)$$

where n is the total number of selected MS events. $p_i = \frac{t_{i+1} - t_i}{t_n - t_1}$, t_i is the occurrence time of the i -th MS event, and the value of p_i is 0~1.

The spatial distribution of MS events is important for understanding the stability of the coal rock mass in the mine.

$$Q_D = \text{Var}(R_n) / \Delta \bar{R}_n \quad (14)$$

where Q_D , $\text{Var}(R_n)$, and $\Delta \bar{R}_n$ are the space concentration, variance, and mean value of the radius corresponding to the last n MS events.

The spatiotemporal diffusion is summarized to reflect the dispersion degree of MS events in time and space.

$$d_s = (\bar{X})^2 / \bar{t} \quad (15)$$

where \bar{X} is the average distance between sequential MS events. \bar{t} is the average time interval between sequential MS events.

The energy concentration index is established to reflect the energy change and MS distribution before the rockburst.

$$Q_E = \text{Var}(E_n) / \Delta \bar{E}_n \quad (16)$$

where Q_E , $\text{Var}(E_n)$, and $\Delta \bar{E}_n$ are the energy concentration, the variance, and the mean value of the energy corresponding to the continuous n MS events.

4.2 The prototype-based rockburst risk and type prediction head

To control the inter-class distance and intra-class variance of the prediction vectors, a novel rockburst risk and type prediction head that can constantly adjust the position of the prototype is proposed, as seen in Figure 4. In the perceptron-based prediction head, the prototype can be regarded as the representative vector in the traditional prediction head.

In this method, the sub-centers of vectors belonging to class C are described by K prototypes, that is, $\{p_{c,k}\}_{k=1}^K$. $p_{c,k}$ is the k -th sub-cluster center of prediction vectors belonging to class C . With this prototype-based classifier, the probability distribution of prediction vector v over the C class can be described as

$$p(c|v) = \frac{\exp(-s_{v,c})}{\sum_{c'=1}^C \exp(-s_{v,c'})}, \text{ with } s_{v,c} = \min \left\{ \langle v, p_{c,k} \rangle \right\}_{k=1}^K \quad (17)$$

where the vector-class distance $s_{v,c} \in [-1, 1]$ is the distance to the closest prototype of class c . Based on Equation 12, the cross-entropy loss is

$$l_{CE} = -\log \frac{\exp(-s_{v,c})}{\sum_{c'=1}^C \exp(-s_{v,c'})} \quad (18)$$

As studied in Section 3, the prediction head should push the prediction vector close to a certain prototype (i.e., the center of vectors) of class c and distant from other prototypes belonging to other classes. However, Equation 7 only considers one vector-class distance. Therefore, the updating method and other limiting conditions of the prototype should be studied.

The prototypes are selected and assigned by the online clustering. Prediction vectors within the same class are assigned to the prototypes of that class, and the prototypes are then updated according to the assignments. Formally, given the vectors $V_c = \{v_n\}_{n=1}^N$ in a training batch that belongs to class c , the goal is to map the vectors V_c to the K prototypes $P_c = \{p_{c,k}\}_{k=1}^K$ of class C . The vector-to-prototype mapping is denoted as $L_c = [l_{v_n,k}]_{n=1}^N \in \{0, 1\}^{K \times N}$. $l_{v_n,k} = [l_{v_n,k}]_{k=1}^K \in \{0, 1\}^K$ is the one-hot assignments of vectors v_n over the K prototypes. The optimization of L_c is achieved by maximizing the similarity between vector embeddings and the prototypes.

$$\begin{aligned} & \max_{L_c} \text{Tr}(L_c^T P_c^T V_c), \\ & \text{s.t. } L_c \in \{0, 1\}^{K \times N}, L_c^T 1^K = 1^N, L_c 1^N = \frac{N}{K} 1^K \end{aligned} \quad (19)$$

The unique assignment constraint $L_c^T 1^K = 1^N$ ensures that each vector is assigned to one and only one prototype. The equipartition

constraint $L_c \mathbf{1}^N = \frac{N}{K} \mathbf{1}^K$ enforces that each prototype is selected at least N/K times in the batch on average. This constraint prevents all vectors from being assigned to a single prototype and eventually benefits the representative ability of the prototypes. To solve Equation 8, L_c can be relaxed to an element of the transportation polytope:

$$\begin{aligned} \max_{L_c} \text{Tr}(L_c^\top P_c^\top V_c) + \kappa h(L_c), \\ \text{s.t. } L_c \in \mathbb{R}_+^{K \times N}, L_c^\top \mathbf{1}^K = \mathbf{1}^N, L_c \mathbf{1}^N = \frac{N}{K} \mathbf{1}^K \end{aligned} \quad (20)$$

where $h(L_c) = \sum_{n,k} l_{v_n,k} \log l_{v_n,k}$ is an entropy, and $\kappa > 0$ is a parameter that controls the smoothness of the distribution. With the soft assignment relaxation and the extra regularization term $h(L_c)$, the solution of Equation 9 can be given as

$$L_c = \text{diag}(\psi) \exp\left(\frac{P_c^\top V_c}{\kappa}\right) \text{diag}(\epsilon) \quad (21)$$

where $\psi \in \mathbb{R}^K$ and $\epsilon \in \mathbb{R}^N$ are re-normalization vectors, computed by a few steps of the Sinkhorn-Knopp iteration.

The vector-prototype contrastive learning strategy is employed to maximize the prototype assignment posterior probability and the variation of different prototypes. The vector-prototype contrastive loss is

$$l_{PPC} = -\log \frac{\exp(v^\top p_{c_v,k_v}/\tau)}{\exp(v^\top p_{c_v,k_v}/\tau) + \sum_{p^- \in P^-} \exp(v^\top p^-/\tau)} \quad (22)$$

where the temperature τ controls the concentration level of representations. Equation 11 enforces each vector v to be similar with its assigned prototype p_{c_v,k_v} , and dissimilar with other irrelevant prototypes p^- .

Equation 18 and Equation 22 inspire inter-class and inter-cluster discrimination but do not consider reducing the intra-cluster variation, that is, making vectors of the same prototype compact. Thus, a compactness-aware loss is employed for further regularizing representations by directly minimizing the distance between each prediction vector and its assigned prototype:

$$l_{PPD} = (1 - v^\top p_{c_v,k_v})^2 \quad (23)$$

Note that both v and p_{c_v,k_v} are 2-normalized. To accurately control the results, the prototypes were not learned by stochastic gradient descent. After each training iteration, each prototype is updated as:

$$p_{c,k} \leftarrow \mu p_{c,k} + (1 - \mu) \bar{v}_{c,k} \quad (24)$$

where the updated weight μ is 0.999. $\bar{v}_{c,k}$ is the mean vector of the embedded training pixels, which are assigned to prototype $p_{c,k}$ by online clustering.

This training objective minimizes intra-cluster variations while maintaining separation between features with different prototype assignments. As shown in Figure 5, the total loss for prototype-based prediction head is

$$L = \lambda_1 l_{CE} + \lambda_2 l_{PPC} + \lambda_3 l_{PPD} \quad (25)$$

where $\lambda_1 \sim \lambda_3$ are the weights for the loss function. Their values are 1, 0.5, and 0.1 in the article.

As shown in Figure 1, there are two outputs of the prototype-based prediction head, that is, the rockburst types and rockburst risks. There are four classes for the rockburst types, including geological structure, hard rock strata, coal pillars, and mining depth. The class of rockburst risks consists of dangerous and non-dangerous events.

5 Engineering application and experiment implementation details

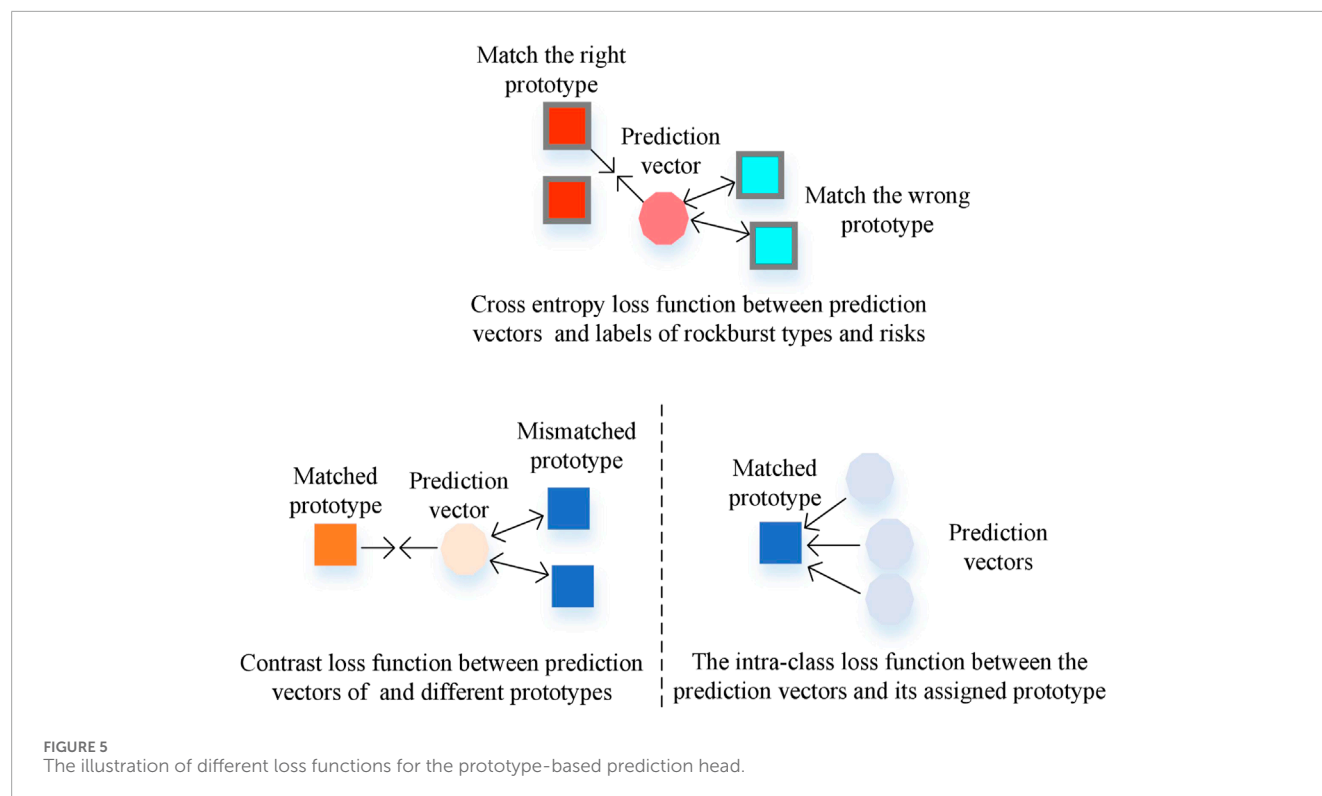
5.1 Rockburst type and risk prediction data

In the experiments, microseismic data from three different rockburst-prone mines are utilized, with the number of monitoring data elements being 29,156, 48,318, and 32,796, respectively. The geological formations and mining conditions exhibit significant heterogeneity for these three mines. Because these mines are located in different places, the geological and mining conditions vary among these rockburst mines. The MS data information includes time interval, X, Y, and Z coordinates, and energy value, respectively. The types of rockburst events and MS events are determined by the projection results of events in different rockburst controlling factor areas. The rockburst types are geological structure, hard rock strata, coal pillars, and mining depth. The rockburst risk levels are dangerous and non-dangerous. In the original MS data, the proportion of rockburst risk events is very low. Therefore, in the training and testing stages, large energy events, obvious mine earthquakes, and rockbursts are defined as dangerous events, and other MS events are defined as non-dangerous samples. In the presented prototype-based prediction head, the prototype numbers for each rockburst type and risk level are set as 5.

In this paper, the standardized time interval of two adjacent events, the X, Y, and Z coordinates, and the energy are selected as the input features of the encoder. The standardized rockburst prediction indexes are selected as the input feature of the decoder. The target is future MS event types and risk levels. By employing the inputs as features and future MS event types and rockburst risk levels as labels, the samples for rockburst types and levels prediction are constructed. When processing the original MS data to sample features, by setting the fused number of MS events as $n = 5, 6, \dots, 15$, thirty datasets for each coal mine are constructed. The partition ratio of training and test samples is 7:3 based on the continuous timeline. Considering the imbalance between different risk level events as well as the number of dangerous and non-dangerous samples, the weighted sampling method is used during training.

5.2 Comparison methods and implementation details

The comparison methods are SVM (Ji et al., 2020), CNN (Zhang et al., 2021), LSTM (Di et al., 2023b), and CNN-GRU with traditional prediction head methods, which are the most popular supervised machine learning methods for rockburst risk and type prediction. The traditional prediction head is a two-layer MLP network.



The training epoch for each method is 200. For the comparison method with the traditional prediction head, the CE loss function is employed to evaluate the difference between the model output and the targets. The input of comparison methods is the combination of sample features. All experiments are implemented on Pytorch 1.10.2 + CUDA 11.3, in FP32 precision by using two RTX A6000 GPUs. A batch size of 512, an Adam optimizer with a momentum of 0.9, a weight decay of 4×10^{-5} , and an initial learning rate of 5×10^{-4} are employed.

To show the performance of different algorithms, the evaluation metric of different algorithms adopts the prediction accuracy of different rockburst types (Types), non-dangerous samples (Non), and dangerous samples (Dan) on the testing dataset.

5.3 Comparison results of different rockburst types and risk prediction methods

The comparison study results are shown in Figure 6. The presented CNN-GRU model with the prototype-based prediction head shows the best performance. The accuracy of the proposed method on the testing set with different rockburst type samples, only non-dangerous samples, only dangerous events achieves 80.72%, 78.62%, and 82.57% in Mine 1, 78.19%, 89.13%, and 79.72% in Mine 2, and 76.07%, 86.92%, and 78.37% in Mine 3.

When different models reach the highest accuracy on different data sets, the corresponding fused MS event number n is similar. For Mine 1 and Mine 2, when the fused MS event number n is 8, almost all methods achieve the best performance. For Mine 3, the best

performance of the entire method appears when the fused MS event number n is 7. This may be because the rockburst prediction indexes and MS data are mainly physical quantity indexes, considering the physical logic of coal rock mass failure and mining rate. For the same mine or workplace, the physical logic and mining rate are similar.

6 Ablation experiments and discussion

6.1 The number of prototypes

The number of prototypes in the prediction head influences the accurate description of the intra-class variance and inter-class distance of the rockburst prediction vectors. It can also affect the output distribution of the prediction vector through the loss function. The results of setting the fused MS event number n to 8/8/7 for Mine 1, Mine 2, and Mine 3 are shown in Table 1. The optimal number of prototypes per class is five, considering the trade-off between performance and computation cost. When the number of prototypes is too small, it is impossible to describe the variance between vectors in different classes. Too many prototypes can result in overfitting and reduced computational efficiency.

6.2 Different deep learning methods with the prototype-based prediction head

The prototype-based rockburst type and risk prediction head showed significant performance gain for the CNN-GRU method, as shown in Figure 6. To demonstrate the universality and superiority

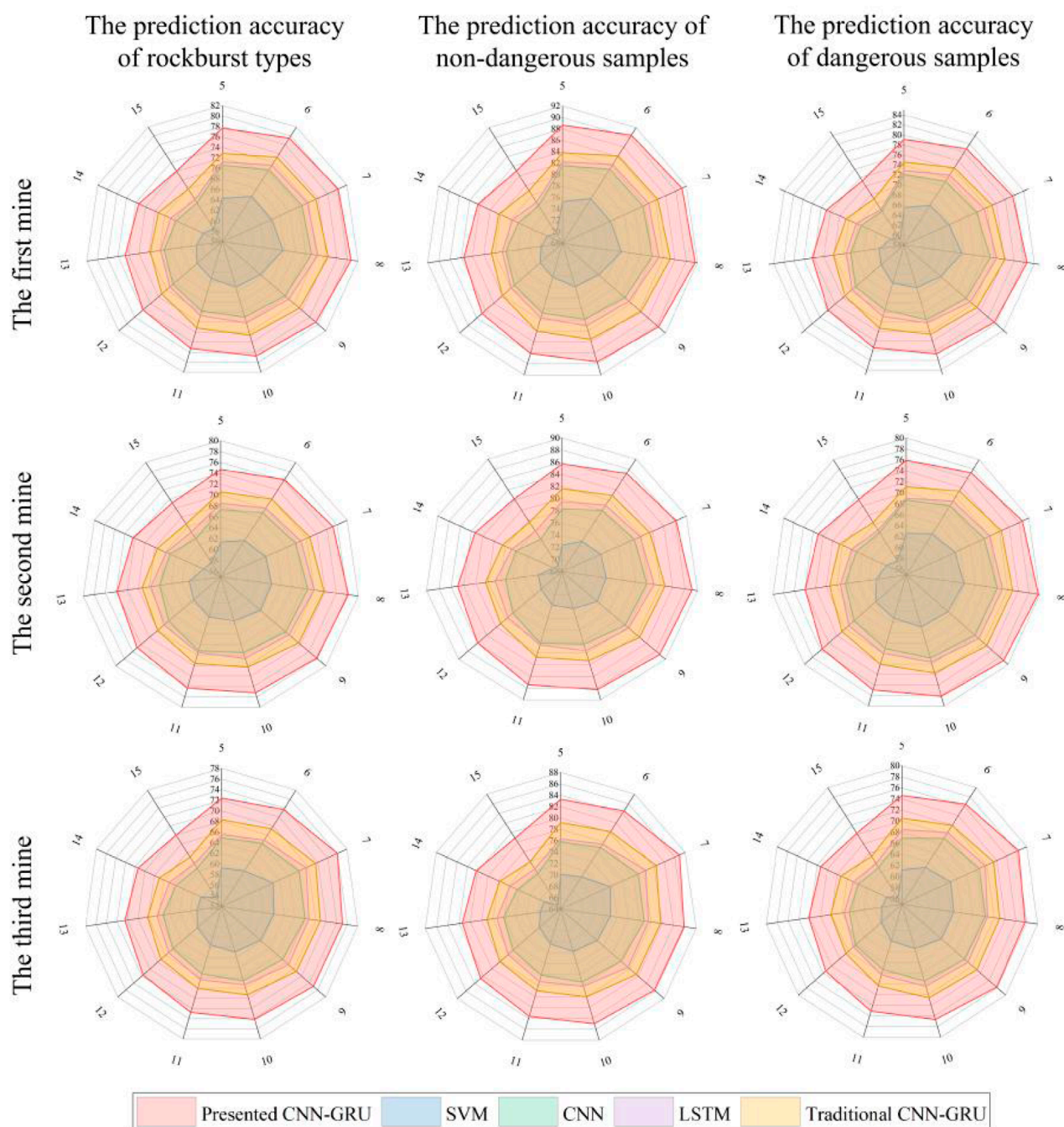


FIGURE 6

The performance comparison of the presented method and other methods. The presented CNN-GRU method is the presented CNN-GRU model with the prototype-based prediction head. The traditional CNN-GRU method is the CNN-GRU model with a two-layer MLP prediction head.

of the prototype-based rockburst type and risk prediction head, the prototype-based prediction head is employed with other methods. The fused MS event number n is set as 8/8/7 for Mine 1, Mine 2, and Mine 3. Table 2 reports that the prototype-based prediction head is obviously superior to the traditional prediction head for the employed deep learning methods. This is mainly because the traditional prediction head only uses the cross-entropy loss function to train, which cannot prevent the

representative vectors of the same class from becoming similar in the training process. This results in the prediction head being unable to describe the inner-class variance even with a sufficient number of representative vectors. At the same time, it is impossible to control the encoder-decoder to ensure the same kind of output prediction vectors close in the distribution space for the traditional prediction head. These problems are overcome in the prototype-based prediction head.

TABLE 1 The influence of prototype number on the presented CNN-GRU model with a prototype-based prediction head.

Number	Accuracy in Mine 1			Accuracy in Mine 2			Accuracy in Mine 3		
	Types	Non	Dan	Types	Non	Dan	Types	Non	Dan
1	77.89	87.69	76.88	75.23	85.03	74.36	74.10	82.77	71.11
3	80.86	90.06	78.92	78.22	87.67	76.60	76.65	85.02	74.00
5	82.57	91.62	80.72	79.72	89.13	78.19	78.37	86.92	76.07
7	82.70	91.82	80.84	79.88	89.29	78.38	78.48	87.08	76.25
10	82.88	91.93	81.04	80.07	89.47	78.48	78.59	87.26	76.40

TABLE 2 The influence of prototype number on the presented CNN-GRU model with the prototype-based prediction head. CNN-GRU-Prototype and CNN-GRU-Tradition are the CNN-GRU model with a prototype-based prediction head and a traditional prediction head, respectively.

Prediction head	Throughput (samples/s)	Accuracy in Mine 1			Accuracy in Mine 2			Accuracy in Mine 3		
		Types	Non	Dan	Types	Non	Dan	Types	Non	Dan
CNN-Prototype	420.8	77.81	86.80	75.95	74.27	84.68	73.07	72.77	81.02	70.15
CNN-Tradition	429.7	75.23	83.99	73.11	72.07	81.80	70.83	70.52	78.94	68.10
LSTM-Prototype	453.6	79.87	88.43	77.40	75.76	85.94	74.97	74.21	83.16	72.34
LSTM-Tradition	462.1	76.48	85.18	74.31	73.05	82.97	71.98	71.70	80.44	69.57
CNN-GRU-Prototype	405.9	82.57	91.62	80.72	79.72	89.13	78.19	78.37	86.92	76.07
CNN-GRU-Tradition	393.3	78.16	87.19	76.29	74.96	84.76	73.78	73.38	82.44	71.53

TABLE 3 The influence of loss function weights on the presented CNN-GRU model. The employed model is CNN-GRU-Prototype. $\lambda_1 \sim \lambda_3$ are the loss function weights for cross-entropy loss, vector-prototype contrastive loss, and loss compactness-aware loss in Equation 25.

Loss function weights			Accuracy in Mine 1			Accuracy in Mine 2			Accuracy in Mine 3		
λ_1	λ_2	λ_3	Types	Non	Dan	Types	Non	Dan	Types	Non	Dan
1	0.5	0.1	82.57	91.62	80.72	79.72	89.13	78.19	78.37	86.92	76.07
1	0	0.1	80.13	88.46	78.12	76.23	86.79	75.48	76.89	84.26	73.82
1	0.5	0	81.27	90.12	78.94	77.53	87.05	77.56	77.24	85.83	75.06
1	0.5	0.5	81.92	90.53	79.64	78.96	88.06	77.29	77.10	85.82	74.95
1	1	0.1	82.01	90.91	80.15	79.36	88.23	77.59	77.63	86.01	75.26

6.3 The weight of loss functions

There are three loss functions in Equation 25. Ablation studies conducted to evaluate their individual contributions, as shown in Table 3, demonstrate that all three components enhance model performance, albeit with varying impacts. The cross-entropy loss exhibits the most significant contribution, as it is the primary training objective. Meanwhile, the vector-prototype contrastive loss and the loss compactness-aware loss primarily regulate prototype distribution within the classifier. Notably,

the experimental results reveal that maximizing the posterior probability of prototype assignments and enhancing prototype diversity between clusters are more critical than minimizing intra-cluster compactness.

6.4 The weight of loss functions

To evaluate the influence of input indicators on model performance, we conducted ablation experiments on five indicators

TABLE 4 The influence of input indexes on the presented CNN-GRU model.

The missing index	Accuracy in Mine 1			Accuracy in Mine 2			Accuracy in Mine 3		
	Types	Non	Dan	Types	Non	Dan	Types	Non	Dan
Temporal concentration	60.21	69.21	58.20	57.23	66.46	54.74	57.10	62.71	53.60
Time information entropy	62.23	71.96	60.86	59.65	68.28	56.77	59.30	64.72	56.03
Space concentration	68.88	77.61	67.04	66.38	74.70	63.84	65.10	72.07	62.55
Spatiotemporal diffusion	72.61	81.50	71.33	69.94	79.14	67.93	68.84	76.51	66.38
Energy concentration index	77.38	86.28	75.40	74.67	83.53	72.23	73.31	81.23	70.90

input to the GRU model. The results are presented in Table 4. The experiments revealed that the energy concentration index contributed most significantly to model performance, while the time concentration index had the least impact. This phenomenon may be attributed to the fact that the energy concentration index directly reflects the accumulation and release processes of elastic strain energy within the rock mass. When energy becomes highly concentrated in a localized area, it indicates sufficient accumulation of strain energy, which may suddenly release upon reaching critical conditions, potentially triggering a rockburst.

7 Conclusion

To overcome the problems of previous rockburst prediction tasks, this paper constructs quantitative information models of geology and mining based on typical rockburst events. By studying the relationship between the spatial-temporal distribution characteristics of MS data before a rockburst and the main controlling factors of rockbursts, the results show that the distribution features may be different for the same type of MS and rockburst events. Different types of events may show similar distribution features. The quantitative research results on the relationship between the deep learning prediction algorithm performance and prediction vectors show that the rockburst prediction model must accurately distinguish among the various types of rockburst events and also accurately describe the intra-class variance of monitoring data.

Based on the above insights, a novel rockburst risk and type prediction algorithm based on a CNN-GRU model with prototype-based prediction is proposed. The CNN-GRU model consists of an encoder and a decoder, which can produce prediction vectors by fusing implicit and explicit information extracted from the original MS data and prediction indexes. The prototype-based prediction uses cross-entropy loss, vector-prototype contrastive loss, and vector-prototype contrastive loss to automatically control the intra-class variance and inter-class gap of rockburst risk and type prediction vectors. The performance superiority of the proposed algorithm compared with the previous algorithm is verified by the comparison

experiment on the data of three mines. The ablation experiment also proves the universality of the proposed prototype-based prediction head in different algorithms for rockburst risk and type prediction.

Data availability statement

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

Author contributions

XZ: Writing – original draft, Writing – review and editing. GL: Writing – review and editing, Writing – original draft. YC: Writing – original draft, Writing – review and editing. HW: Writing – review and editing, Writing – original draft. HZ: Writing – original draft, Writing – review and editing. HL: Writing – original draft, Writing – review and editing. WD: Conceptualization, Visualization, Investigation, Software, Validation, Funding acquisition, Formal analysis, Supervision, Writing – review and editing, Resources, Data curation, Methodology, Project administration, Writing – original draft. XL: Writing – original draft, Writing – review and editing. XX: Writing – original draft, Writing – review and editing. YH: Writing – review and editing, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The authors would like to acknowledge the support of Chinese Institute of Coal Science, Beijing 100013, China; the Shandong Energy Group Co., LTD, Jinan Shandong 250014, China (No. SNKJ2023A17-R01, No. SNKJ2022BJ03-R28); the Taishan Industrial Experts Program (No. tscx202408130); and the Key Science and Technology Project of the Ministry of Emergency Management of the People's Republic of China (2024EMST070702).

Conflict of interest

Authors XZ, GL, YC, and HW were employed by Shandong Energy Group Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The authors declare that this study received funding from Shandong Energy Group. The funder had the following involvement in the study: data collection and analysis, decision to publish, and preparation of the manuscript.

References

- Adoko, A., and Zvarivadza, T. (2018). A bayesian approach for predicting rockburst in ARMA US Rock Mechanics/Geomechanics Symposium. Seattle, Washington ARMA.
- AskariPour, M., Saeidi, A., Rouleau, A., and Mercier-Langevin, P. (2022). Rockburst in underground excavations: a review of mechanism, classification, and prediction methods. *Undergr. Space* 7 (4), 577–607. doi:10.1016/j.undsp.2021.11.008
- Aydan, Ö., Geniş, M., Akagi, T., and Kawamoto, T. (2017). *Assessment of susceptibility of rock bursting in tunnelling in hard rocks, Modern tunneling science and technology*. London: Routledge, 391–396.
- Basnet, P. M. S., Mahtab, S., and Jin, A. (2023). A comprehensive review of intelligent machine learning based predicting methods in long-term and short-term rock burst prediction. *Tunn. Undergr. Space Technol.* 142, 105434. doi:10.1016/j.tust.2023.105434
- Chen, F., Liang, Z., and Cao, A. (2023). ConvLSTM for predicting short-term spatiotemporal distribution of seismic risk induced by large-scale coal mining. *Nat. Resour. Res.* 32 (3), 1459–1479. doi:10.1007/s11053-023-10193-5
- Cheng, S., Yin, X., Gao, F., and Pan, Y. (2024). Microseismic data-driven short-term rockburst evaluation in underground engineering with strategic data augmentation and extremely randomized forest. *Mathematics* 12 (22), 3502. doi:10.3390/math12223502
- Cheng, T., He, M., Li, H., Liu, D., Qiao, Y., and Hu, J. (2023). Experimental investigation on the influence of a single structural plane on rockburst. *Tunn. Undergr. Space Technol.* 132, 104914. doi:10.1016/j.tust.2022.104914
- Di, Y., Wang, E., Li, Z., Liu, X., Huang, T., and Yao, J. (2023a). Predicting microseismic, acoustic emission and electromagnetic radiation data using neural networks. *J. Rock Mech. Geotechnical Eng.* 16, 616–629. doi:10.1016/j.jrmge.2023.05.012
- Di, Y., Wang, E., Li, Z., Liu, X., Huang, T., and Yao, J. (2023b). Comprehensive early warning method of microseismic, acoustic emission, and electromagnetic radiation signals of rock burst based on deep learning. *Int. J. Rock Mech. Min. Sci.* 170, 105519. doi:10.1016/j.ijrmms.2023.105519
- Dong, L., Shu, H., Tang, Z., and Yan, X. (2023). Microseismic event waveform classification using CNN-based transfer learning models. *Int. J. Min. Sci. Technol.* 33, 1203–1216. doi:10.1016/j.ijmst.2023.09.003
- He, M., Cheng, T., Qiao, Y., and Li, H. (2023). A review of rockburst: experiments, theories, and simulations. *J. Rock Mech. Geotechnical Eng.* 15 (5), 1312–1353. doi:10.1016/j.jrmge.2022.07.014
- He, M., Ren, F., and Liu, D. (2018). Rockburst mechanism research and its control. *Int. J. Min. Sci. Technol.* 28 (5), 829–837. doi:10.1016/j.ijmst.2018.09.002
- Hu, L., Feng, X.-T., Yao, Z.-B., Zhang, W., Niu, W.-J., Bi, X., et al. (2023). Rockburst time warning method with blasting cycle as the unit based on microseismic information time series: a case study. *Bull. Eng. Geol. Environ.* 82 (4), 121. doi:10.1007/s10064-023-03141-3
- Ji, B., Xie, F., Wang, X., He, S., and Song, D. (2020). Investigate contribution of multi-microseismic data to rockburst risk prediction using support vector machine with genetic algorithm. *IEEE Access* 8, 58817–58828. doi:10.1109/access.2020.2982366
- Jin, A., Basnet, P. M. S., and Mahtab, S. (2022). Microseismicity-based short-term rockburst prediction using non-linear support vector machine. *Acta Geophys.* 70 (4), 1717–1736. doi:10.1007/s11600-022-00817-4
- Jinqiang, W., Basnet, P., and Mahtab, S. (2021). Review of machine learning and deep learning application in mine microseismic event classification. *Mining Mineral Deposits* 15, 19–26. doi:10.33271/mining15.01.019
- Li, N., Huang, B., Zhang, X., Yuyang, T., and Li, B. (2019). Characteristics of microseismic waveforms induced by hydraulic fracturing in coal seam for

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

coal rock dynamic disasters prevention. *Saf. Sci.* 115, 188–198. doi:10.1016/j.ssci.2019.01.024

Li, N., Zare Naghadehi, M., and Jimenez, R. (2020). Evaluating short-term rock burst damage in underground mines using a systems approach. *Int. J. Min. Reclam. Environ.* 34 (8), 531–561. doi:10.1080/17480930.2019.1657654

Liang, W., Sari, A., Zhao, G., McKinnon, S. D., and Wu, H. (2020). Short-term rockburst risk prediction using ensemble learning methods. *Nat. Hazards* 104, 1923–1946. doi:10.1007/s11069-020-04255-7

Liang, W., Sari, Y. A., Zhao, G., McKinnon, S. D., and Wu, H. (2021). Probability estimates of short-term rockburst risk with ensemble classifiers. *Rock Mech. Rock Eng.* 54, 1799–1814. doi:10.1007/s00603-021-02369-3

Liang, W., and Zhao, G. (2022). A review of long-term and short-term rockburst risk evaluations in deep hard rock. *J. Rock Mech. Eng.* 41, 19–39. doi:10.13722/j.cnki.jrme.2021.0165

Ma, J., Ma, C., Jun, Z., Li, T., and Wang, X. (2021). Research on prediction of rockburst microseismic parameters based on CNN-LSTM hybrid model. *IOP Conf. Ser. Earth Environ. Sci.* 861, 052097. doi:10.1088/1755-1315/861/5/052097

Manouchehrian, A., and Cai, M. (2018). Numerical modeling of rockburst near fault zones in deep tunnels. *Tunn. Undergr. Space Technol.* 80, 164–180. doi:10.1016/j.tust.2018.06.015

Pu, Y., Apel, D. B., Liu, V., and Mitri, H. (2019). Machine learning methods for rockburst prediction-state-of-the-art review. *Int. J. Min. Sci. Technol.* 29 (4), 565–570. doi:10.1016/j.ijmst.2019.06.009

Wang, J., Apel, D. B., Pu, Y., Hall, R., Wei, C., and Sepehri, M. (2021). Numerical modeling for rockbursts: a state-of-the-art review. *J. Rock Mech. Geotechnical Eng.* 13 (2), 457–478. doi:10.1016/j.jrmge.2020.09.011

Xue, Y., Li, Z., Song, D., He, X., Wang, H., Zhou, C., et al. (2023). A method to predict rockburst using temporal trend test and its application. *J. Rock Mech. Geotechnical Eng.* 16, 909–923. doi:10.1016/j.jrmge.2023.07.017

Yang, J., Chen, W., Tan, X., and Yang, D. (2018). Analytical estimation of stress distribution in interbedded layers and its implication to rockburst in strong layer. *Tunn. Undergr. Space Technol.* 81, 289–295. doi:10.1016/j.tust.2018.07.007

Yin, X., Cheng, S., Yu, H., Pan, Y., Liu, Q., Huang, X., et al. (2024b). Probabilistic assessment of rockburst risk in TBM-excavated tunnels with multi-source data fusion. *Tunn. Undergr. Space Technol.* 152, 105915. doi:10.1016/j.tust.2024.105915

Yin, X., Liu, Q., Huang, X., and Pan, Y. (2021a). Real-time prediction of rockburst intensity using an integrated CNN-Adam-BO algorithm based on microseismic data and its engineering application. *Tunn. Undergr. Space Technol.* 117, 104133. doi:10.1016/j.tust.2021.104133

Yin, X., Liu, Q., Lei, J., Pan, Y., Huang, X., and Lei, Y. (2024a). Hybrid deep learning-based identification of microseismic events in TBM tunnelling. *Measurement* 238, 115381. doi:10.1016/j.measurement.2024.115381

Yin, X., Liu, Q., Pan, Y., and Huang, X. (2021b). A novel tree-based algorithm for real-time prediction of rockburst risk using field microseismic monitoring. *Environ. Earth Sci.* 80, 504–519. doi:10.1007/s12665-021-09802-4

Yin, X., Liu, Q., Pan, Y., Huang, X., Wu, J., and Wang, X. (2021c). Strength of stacking technique of ensemble learning in rockburst prediction with imbalanced data: comparison of eight single and ensemble models. *Nat. Resour. Res.* 30, 1795–1815. doi:10.1007/s11053-020-09787-0

Zhang, H., Zeng, J., Ma, J., Fang, Y., Ma, C., Yao, Z., et al. (2021). Time series prediction of microseismic multi-parameter related to rockburst based on deep learning. *Rock Mech. Rock Eng.* 54 (12), 6299–6321. doi:10.1007/s00603-021-02614-9