

Article

Signal Super Prediction and Rock Burst Precursor Recognition Framework Based on Guided Diffusion Model with Transformer

Mingyue Weng ¹, Zinan Du ^{2,*}, Chuncheng Cai ³, Enyuan Wang ^{2,4,*}, Huilin Jia ^{2,4} , Xiaofei Liu ^{2,4} , Jinze Wu ², Guorui Su ⁵ and Yong Liu ⁶

¹ Shanghai Datun Energy Resources Co., Ltd., Shanghai 200120, China; yjlmkwm@sina.com

² School of Safety Engineering, China University of Mining and Technology, Xuzhou 221116, China; hl.j@cumt.edu.cn (H.J.); liuxiaofei@cumt.edu.cn (X.L.); tb23120010a41ld@cumt.edu.cn (J.W.)

³ Kongzhuang Coal Mine, Shanghai Datun Energy Resources Co., Ltd., Xuzhou 221600, China; chechuncheng1986@126.com

⁴ Key Laboratory of Gas and Fire Control for Coal Mines, China University of Mining and Technology, Xuzhou 221116, China

⁵ Information Institution of Ministry of Emergency Management, Beijing 100029, China; suguorui1991@163.com

⁶ Ministry of Science, Technology and Environmental Protection, Shanghai Datun Energy Resources Co., Ltd., Shanghai 200120, China; liuyong8@chinacoal.com

* Correspondence: duzinan@cumt.edu.cn (Z.D.); weytong@cumt.edu.cn (E.W.)

Abstract: Implementing precise and advanced early warning systems for rock bursts is a crucial approach to maintaining safety during coal mining operations. At present, FEMR data play a key role in monitoring and providing early warnings for rock bursts. Nevertheless, conventional early warning systems are associated with certain limitations, such as a short early warning time and low accuracy of early warning. To enhance the timeliness of early warnings and bolster the safety of coal mines, a novel early warning model has been developed. In this paper, we present a framework for predicting the FEMR signal in deep future and recognizing the rock burst precursor. The framework involves two models, a guided diffusion model with a transformer for FEMR signal super prediction and an auxiliary model for recognizing the rock burst precursor. The framework was applied to the Buertai database, which was recognized as having a rock burst risk. The results demonstrate that the framework can predict 360 h (15 days) of FEMR signal using only 12 h of known signal. If the duration of known data is compressed by adjusting the CWT window length, it becomes possible to predict data over longer future time spans. Additionally, it achieved a maximum recognition accuracy of 98.07%, which realizes the super prediction of rock burst disaster. These characteristics make our framework an attractive approach for rock burst predicting and early warning.

Keywords: rock burst; FEMR; guided diffusion model; transformer; signal super prediction; precursor recognition



Academic Editor: Nikolaos Koukoulas

Received: 27 January 2025

Revised: 21 February 2025

Accepted: 21 February 2025

Published: 17 March 2025

Citation: Weng, M.; Du, Z.; Cai, C.; Wang, E.; Jia, H.; Liu, X.; Wu, J.; Su, G.; Liu, Y. Signal Super Prediction and Rock Burst Precursor Recognition Framework Based on Guided Diffusion Model with Transformer. *Appl. Sci.* **2025**, *15*, 3264. <https://doi.org/10.3390/app15063264>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Rock burst, a major hazard in coal mining, frequently results in significant casualties and property damage. Investigating risk indicators prior to rock burst events is essential for forecasting such disasters in mining operations; some scholars have studied the current status of coal and gas outburst risk identification in underground coal mines, analyzed the challenges and limitations of existing methods, and proposed directions for future research [1–3]. Therefore, it is critical that we use a method to accurately obtain the precursor characteristics of rock burst as soon as possible. To accurately forecast rock burst disasters,

numerous monitoring and early warning techniques have been developed, achieving significant advancements, including the fracture-induced electromagnetic radiation (FEMR) monitoring method, which are generally applied in the field of rock burst disaster early warning [4,5]. Researchers have investigated the formation mechanisms and transmission properties of FEMR signals; some scholars have provided insights into the activity of the Dead Sea transition zone by studying electromagnetic radiation signals from rock fractures before and after earthquakes in the Syrian–Turkish border area. Others have explored the electromagnetic radiation characteristics of rock fractures near the seismically active area in southern Israel [6–8]. Furthermore, numerous studies have explored the time–frequency domain features relevant to rock burst FEMR monitoring and early warning. Some scholars have proposed an early warning method based on microseismic monitoring for rockburst monitoring in Qinling water tunnels. In addition, some scholars have studied the improved waveform adaptive threshold method for microseismic signal denoising [9–11]. Thus, the FEMR monitoring process offers convenience, ensures continuity in monitoring data, and demonstrates efficacy in responding to rock burst precursory information and early warnings. However, the workload associated with manual procedures for analyzing rock burst precursor characteristics within FEMR signals is substantial, leading to an insufficient advance in warning time and inadequate accuracy in the recognition of precursor characteristics [12,13].

The enhancement in computational power has led to remarkable advancements in diverse deep-learning model algorithms and applications. These developments have enabled breakthroughs in areas such as computer vision, time-series data analysis, and beyond [14–18]. In the domain of geophysical studies, advanced machine-learning methodologies, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have been utilized across various geophysical methods, which can not only realize recognizing the precursor characteristics of rock burst, but also predict the risk coefficient of disasters [19–21]. However, in the above method, the timing advance for early warning through precursor characteristics recognition is limited, generally only within about 7 days in advance. Some scholars have proposed a rock burst precursor electromagnetic radiation signal recognition and early warning method based on recursive neural network (RNN). Simultaneously, if the precursor characteristics signal lasts too long, it can result in delayed warning [5,22]. Hence, some scholars have opted for an alternative methodology, and used the RNNs and the transformer architecture to realize the disaster early warning in the type of time-series prediction. The time-series prediction method can predict FEMR signal data about 7 days in future, achieving the signal prediction in the short-term future. Some scholars have proposed a meteorological early warning method based on deep neural networks for predicting landslides caused by precipitation [23–25]. However, to achieve the high level of signal prediction accuracy, this method necessitates the integration of multi-variable data through multi-source fusion training, demanding an extensive array of sensor types deployed onsite. Simultaneously, each prediction for the short-term future requires a substantial volume of pre-existing data, imposing significant storage demands on coal mine databases, which is a challenge for many coal mines to meet. Some scholars have proposed a method for predicting the occurrence patterns of coal mine gas based on multi-source data fusion [26,27]. Given these constraints, there is a pressing need for a rock burst early warning method capable of predicting an FEMR signal in the deep future with minimal reliance on known univariate signal data and recognizing precursor characteristics.

Generative adversarial networks (GANs) consist of two parts: a generator that produces new data from random noise, and a discriminator that evaluates the generated samples against real data. These networks are trained simultaneously to achieve a Nash equilibrium, where neither network can improve its performance without the other adjust-

ing its approach. GANs are recognized as an unsupervised learning technique capable of generating new data that align with the distribution of the training dataset [28–32]. Some scholars have proposed guided learning methods to improve the training process of GANs. Guided generative adversarial networks (GGANs), an advanced variant of GANs, enable controlled data generation by incorporating auxiliary information like labels. These networks have shown enhanced performance over prediction-based models, particularly when applied to unseen data with limited training samples [33–36]. These networks have been used to solve the spectrogram out-painting task, and predict the loss time-frequency data, which have the ability to predict the missing data [37–39]. While GAN-based architectures present numerous advantages compared to predictive models, including their proficiency in generalizing predictions across smaller and varied datasets and their aptitude for capturing complex conditional distributions, they are not devoid of limitations. A significant concern is the phenomenon of “mode collapse”, wherein the generator yields outputs with restricted diversity, thereby inadequately reflecting the full spectrum of variability inherent in the training data. This issue can hinder the model’s ability to generate realistic and diverse samples, ultimately impacting its effectiveness in applications requiring high variability and fidelity [33,34,40]. Moreover, these networks may encounter the challenge of “non-convergence”, where the learning process fails to achieve a Nash equilibrium, leading to unstable performance. To address these shortcomings of GANs, diffusion-based models have been introduced as an alternative approach [41,42].

To overcome the constraints inherent in GANs, diffusion models have been introduced. These models systematically incorporate Gaussian noise into the training samples until they are fully transformed into pure noise. This process enables the neural network to gradually eliminate noise and extract valuable insights from the data [43–45]. Recently, guided diffusion models have demonstrated remarkable prowess in tasks involving pixel matrix data predictions [46–48]. Notably, they have been successfully applied in generating pixel matrix data from textual descriptions, achieving pixel matrix data resolution and inpainting, and facilitating pixel-to-pixel convert, and pixel matrix data out-painting [49–51]. This underscores the extraordinary potential of diffusion-based models to enhance the accuracy of FEMR signal prediction in the deep future.

The guided diffusion model provides new ideas for signal prediction tasks [52–55]. This study converts the FEMR signal time-series data into corresponding time–frequency data through Fourier transform, and uses the guided diffusion model to predict the FEMR signal time-frequency data in the deep future based on known data. In order to achieve FEMR signal prediction in the deep future, we added the recurrent structure in the guided diffusion model in the super prediction process. However, diffusion models, which focus on predicting the incremental noise introduced during each iteration, face challenges when a fixed amount of noise is applied uniformly across the dataset. This can cause some samples to degrade more rapidly than others, introducing instability during training and resulting in errors in the predicted data outcomes [56–59]. Furthermore, if it involves the continuous generation of signal data in the deep future based on partially known data, the error of each generation and the prediction result of the model will accumulate in the data prediction process, which is fatal for deep future data prediction tasks [60–63]. This requires using a method to restore the data predicted in each recurrence by the guided diffusion model to ensure that it provides accurate known data for the next recurrence.

Over the past five years, deep learning is witnessing a remarkable revival, largely fueled by the emergence and widespread adoption of transformer-based architectures [64–67]. The transformer-based methods have shown impressive performance in pixel matrix data restoration tasks, such as pixel matrix data super-resolution. This study added a transformer-based data restoration architecture in the prediction process, which can restore the missing

pixels in the predicted data, and provide higher quality known data for the next recurrence in the prediction process. To conclude, the transformer-based data restoration architecture significantly contributes to the efficacy of guided diffusion models in predicting deep-future FEMR signal time–frequency data, which can realize the FEMR signal super prediction.

This study introduces a novel guided diffusion model with a transformer-based framework that can predict the FEMR signal in the deep future through a minimal amount of known FEMR signal and an auxiliary model to recognize rock burst precursor characteristics. Our methodology consists of a two-stage process, as illustrated in Figure 1, which outlines the framework’s schematic. Initially, we employ a guided diffusion model integrated with a transformer to forecast FEMR signals over an extended future period. Subsequently, a transformer-based auxiliary model is utilized to identify the precursor characteristics of rock bursts within the FEMR signals. This approach enables the prediction of FEMR signals up to 360 h (15 days) into the future using only 12 h of known signal data. As a result, the time advance for rock burst early warning, as well as the precision of precursor characteristic identification, have been significantly enhanced.

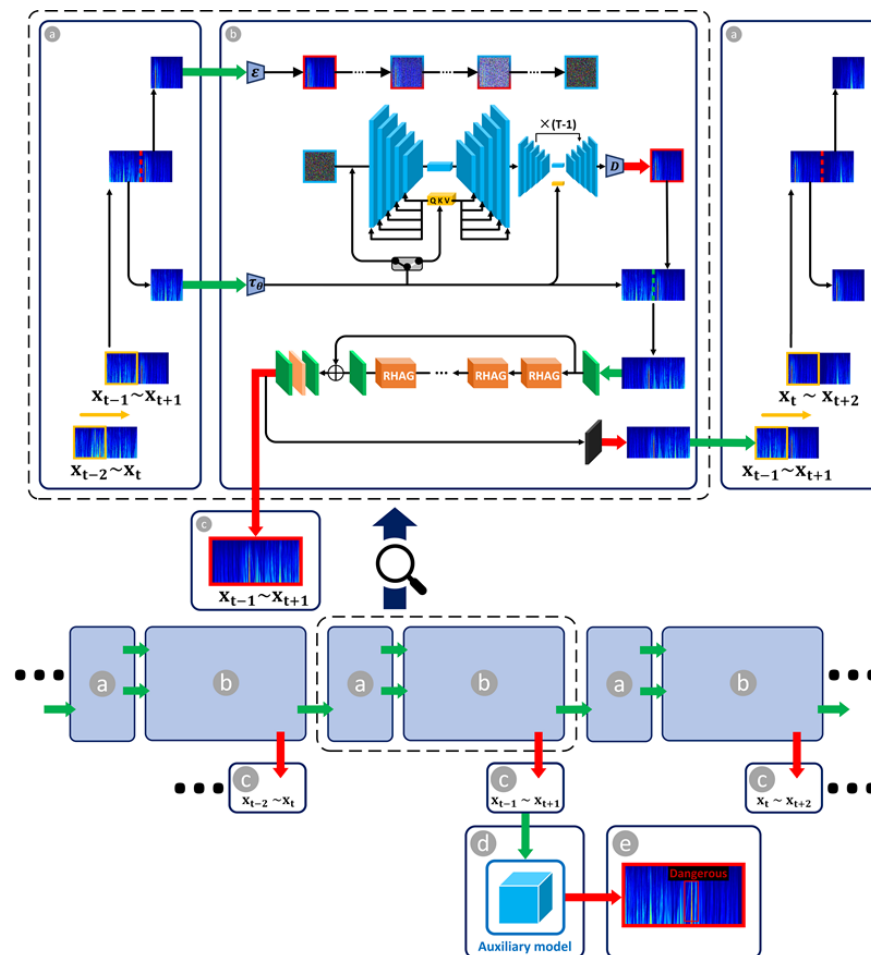


Figure 1. Description of workflow. (a) Window sliding operation: the data range of the window selection slides from $x_{t-2} \sim x_{t-1}$ ($x_{t-1} \sim x_t$) to $x_{t-1} \sim x_t$ ($x_t \sim x_{t+1}$); the data range of the time–frequency data slides from $x_{t-2} \sim x_t$ ($x_{t-1} \sim x_{t+1}$) to $x_{t-1} \sim x_{t+1}$ ($x_t \sim x_{t+2}$). (b) Add Gaussian noise to the ground truth time–frequency data that needs to be predicted; inputs: the known part of the time–frequency data and pure Gaussian noise; network architecture of guided diffusion model with transformer; (c) outputs: time–frequency data in the $x_{t-1} \sim x_{t+1}$ time range predicted by guided diffusion model with transformer; (d) inputs: summary of time–frequency data prediction results of the guided diffusion model with transformer; and (e) outputs: the auxiliary model recognition results of rock burst precursor characteristics.

2. Methodology

2.1. Auxiliary Model

The guided diffusion model with transformer can predict the FEMR signal data in deep future. To achieve effective early warning of rock bursts, it is essential that we identify precursor characteristics within FEMR time–frequency data. Transformers, which rely on self-attention mechanisms, have gained widespread popularity in natural language processing tasks. Recently, there has been a rising focus in adapting transformer-based methods to a variety of computer vision applications [67–70]. In our study, we employ a transformer-based architecture as the core model for identifying precursor characteristics of rock bursts.

Figure 2 illustrates the schematic of the auxiliary model designed to identify precursor characteristics of rock bursts in FEMR time–frequency data. The input data, which consist of summarized FEMR time–frequency data predicted by the guided diffusion model with transformer, is initially processed using a non-overlapping convolution to generate patch embeddings, followed by a normalization layer. The auxiliary model comprises four stages: The first two stages focus on learning local characteristics through attention mechanisms. In the third and fourth stages, the FEMR time–frequency characteristic maps are first processed by a window local attention mechanism to aggregate local information, and then passed through a deformable attention block to model global relationships between the locally enhanced tokens. The output of the auxiliary model provides the recognized precursor characteristics of rock bursts, with “Dangerous” indicators marked by red boxes.

Deformable attention block, as illustrated in Figure 2d, is the key to recognizing the precursor characteristics of rock burst. Given the input characteristics map x of FEMR time–frequency data, a uniform grid of points is generated as the references. To determine the displacement for each reference location, as shown in Equation (1), the feature maps are linearly mapped to the query tokens $q = xW_q$, as shown in Equation (2), and then directed into a compact sub-network $\theta_{offset}(\cdot)$ to output the offsets $\Delta p = \theta_{offset}(q)$. Subsequently, the features are extracted at the positions of the deformed points to serve as keys and values, followed by the application of projection matrices:

$$q = xW_q, \tilde{k} = \tilde{x}W_k, \tilde{v} = \tilde{x}W_v, \quad (1)$$

$$\Delta p = \theta_{offset}(q), \tilde{x} = \mathcal{O}(x; p + \Delta p). \quad (2)$$

where q , $\tilde{k}$, and $\tilde{v}$ denote query, deformed key, and deformed value embeddings, respectively. W_q , W_k , W_v are matrices.

Specifically, as shown in Equation (3), we set the sampling function $\mathcal{O}(\cdot; \cdot)$ to a bilinear interpolation process to ensure differentiability:

$$\mathcal{O}(z; (p_x, p_y)) = \sum_{(r_x, r_y)} g(p_x, r_x)g(p_y, r_y)z[r_y, r_x, :] \quad (3)$$

where $g(a, b) = \max(0, 1 - |a - b|)$ and (r_x, r_y) indexes all the locations on z . Since it is non-zero exclusively at the four nearest integral points to g , it reduces Equation (3) to a weighted average across these four positions.

Following established methodologies, as shown in Equation (4), we apply multi-head attention to q, k, v and incorporate relative positional offsets R . The output of a single attention head is expressed as follows:

$$z^{(m)} = \sigma \left(q^{(m)} \tilde{k}^{(m)\top} / \sqrt{d} + \phi(\hat{B}; R) \right) \tilde{v}^{(m)} \quad (4)$$

where $\sigma(\cdot)$ represents the softmax function, and $d = C/M$ indicates the dimensionality of each head. $z^{(m)}$ refers to the embedding output generated by the m -th attention head. $\phi(\hat{B}; R)$ are derived from the position embedding, following established methodologies but with specific modifications.

The features from each head are combined through W_o concatenation and then transformed to produce the final output z , as described in Equation (5).

$$z = \text{Concat}(z^{(1)}, \dots, z^{(M)}) W_o \quad (5)$$

where W_o are the projection matrices.

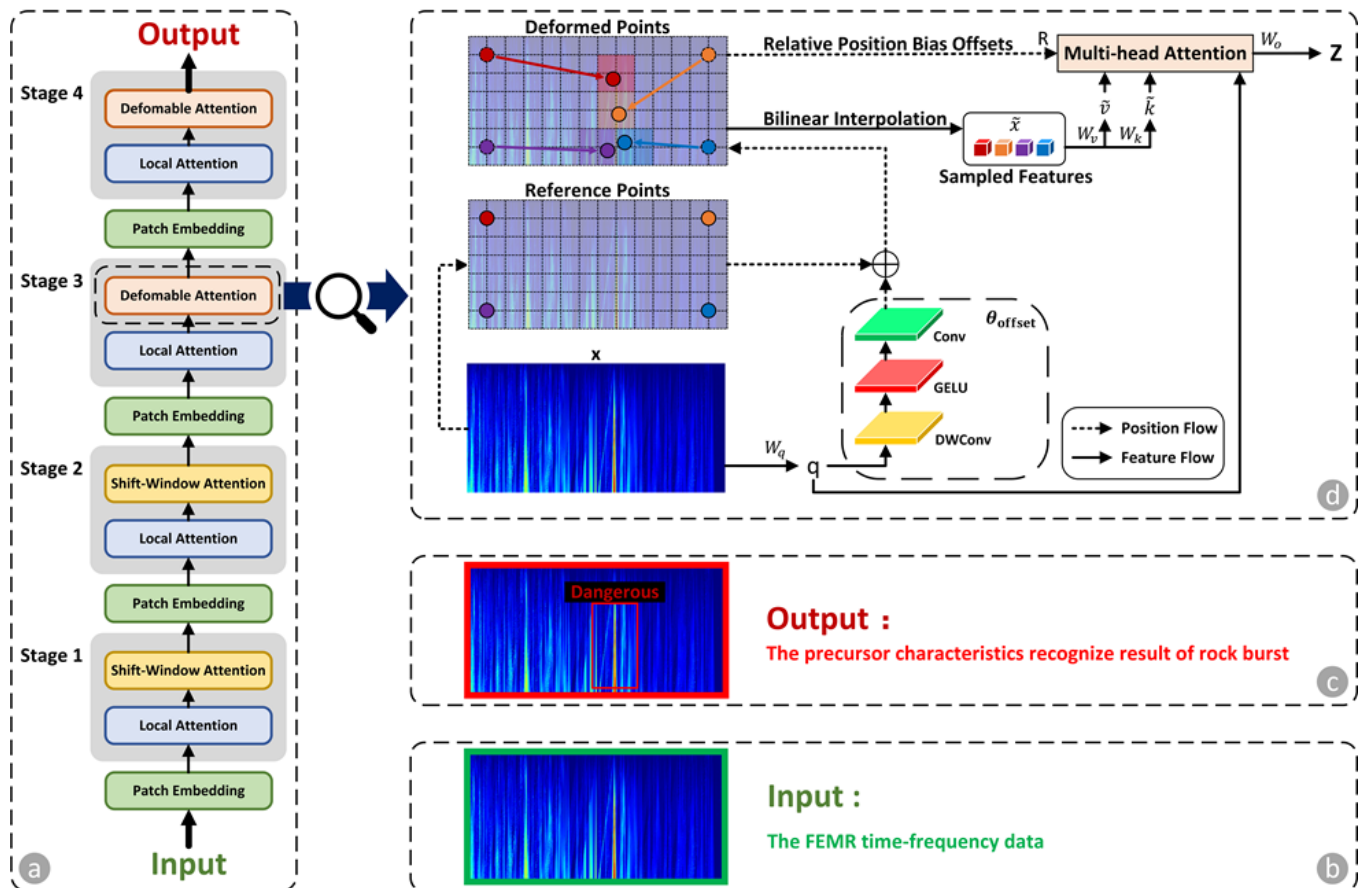


Figure 2. Schematic of auxiliary model with transformer-based architecture backbone: (a) schematic of auxiliary model with transformer-based architecture backbone; (b) inputs: the summary of FEMR time–frequency data predicted by the guided diffusion model with transformer; (c) outputs: the precursor characteristics recognize result of rock burst in FEMR time–frequency data; and (d) deformable attention block.

2.2. Guided Diffusion Model with Transformer

To achieve FEMR signal super prediction, the guided diffusion model with transformer can predict the signal in deep future. In the recurrence structure of the guided diffusion model with transformer, each 12 h future can be predicted through the data from previous 12 h. In order to obtain high-quality future FEMR signal time–frequency data in each recurrence of prediction tasks, we added a transformer-based data restoration architecture in guided diffusion model to restore the missing pixels in predicted FEMR signal data.

2.2.1. Guided Diffusion Model

Diffusion models are probabilistic frameworks that seek to model data distribution by iteratively refine a variable initialized with a normal distribution through a denoising process. This process involves establishing a Markov chain of diffusion steps, where systematic noise is gradually introduced to the input data during the forward diffusion process [71]. Considering the future time–frequency data as input z_0 , as shown in Equations (6) and (7), the forward diffusion process can be represented as a Markov chain with a length of T :

$$q(z_t|z_{t-1}) := N\left(z_t; \sqrt{1 - \beta_t}z_{t-1}, \beta_t I\right); \beta_t \in (0, 1); t \in \{1, \dots, T\} \quad (6)$$

$$q(z_{1:T}|z_0) := \prod_{t=1}^T q(z_t|z_{t-1}); t \in \{1, \dots, T\} \quad (7)$$

where $\beta_t \in (0, 1)$ is the variance scheduler that regulates the step size of the noise added $t \in \{1, \dots, T\}$ is the t uniformly sampled from $\{1, \dots, T\}$.

Our goal is to predict future time–frequency data from noise input $z_t \approx N(0, I)$ given a condition (known data), with the trained guided diffusion model consisting of future data encoder ε , predicted data decoder D , and known data encoder τ_θ , which paves the way to controlling the prediction process via known data y . Given time–frequency data x , the encoder ε encodes x into a latent representation $z = \varepsilon(x)$, and the decoder D reconstructs the time–frequency data from the latent, giving $\tilde{x} = D(z) = D(\varepsilon(x))$.

The data prediction can be achieved using a conditional denoising autoencoder $\epsilon_\theta(z_t, t, \tau_\theta(y))$, the conditional denoising autoencoder can predict the future time–frequency data given known data conditions. With a U-Net backbone and cross-attention mechanism, the guided diffusion model excels at learning attention-based models from known data. To pre-process known data y , we introduce an encoder τ_θ that projects y to an intermediate representation $\tau_\theta(y) \in R^{M \times d_\tau}$, which is then mapped to the intermediate layers of the U-Net through a cross-attention layer that implements the attention $(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right) \cdot V$, with $Q = W_Q^{(i)} \cdot \varphi_i(z_t)$, $K = W_K^{(i)} \cdot \tau_\theta(y)$, and $V = W_V^{(i)} \cdot \tau_\theta(y)$. Here, $\varphi_i(z_t) \in R^{N \times d_\epsilon^i}$ denotes a (flattened) intermediate representation of the U-Net, which is used for implementing ϵ_θ , and $W_V^{(i)} \in R^{d \times d_\epsilon^i}$, $W_Q^{(i)} \in R^{d \times d_\tau}$ & $W_K^{(i)} \in R^{d \times d_\tau}$ are learnable projection matrices. See Figure 3 for a visual depiction. Based on known data-conditioning pairs, as shown in Equation (8), we subsequently learn the conditional L_{SDM} via:

$$L_{SDM} := E_{\varepsilon(x), y, \epsilon \sim N(0,1), t} \left[\|\epsilon - \epsilon_\theta(z_t, t, \tau_\theta(y))\|_2^2; t \in \{1, \dots, T\} \right] \quad (8)$$

where both τ_θ and ϵ_θ are jointly optimized via Equation (8).

2.2.2. Transformer-Based Data Restoration Architecture

As shown in Figure 4, to acquire high-quality future FEMR signal time–frequency data in each iteration of prediction tasks, we integrated a transformer-based data restoration architecture within the guided diffusion model. This addition is specifically designed to restore the missing pixels in the predicted FEMR signal data.

The transformer-based data restoration architecture adopts a residual in residual (RIR) structure, comprising three key components: shallow feature extraction, deep feature extraction, and image reconstruction. The shallow characteristics extraction part uses a 3×3 convolution layer to extract shallow characteristics from the low-quality input image as Equation (9):

$$F_0 = H_{SF}(I_{LQ}) \quad (9)$$

where F_0 is the shallow characteristics and I_{LQ} is low-quality (LQ) input image.

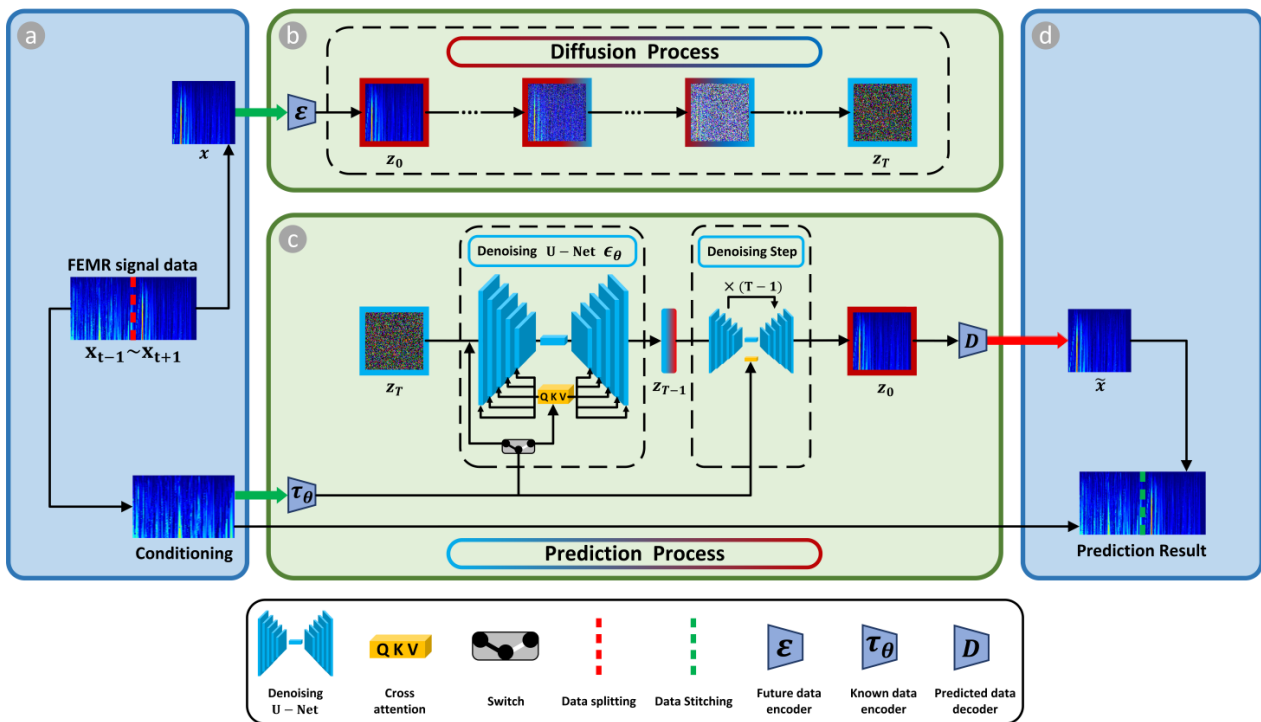


Figure 3. Schematic of guided diffusion model: (a) data splitting: split the FEMR signal data from x_{t-1} to x_{t+1} into conditioning known data from x_{t-1} to x_t and future data x_t to x_{t+1} ; (b) diffusion process: encode the data through the future data encoder ϵ , and tractable noise is added to the input data during the diffusion process; (c) prediction process: the U-Net-based prediction process calculates the characteristics similarity between known and future data, and the trained model can predict future data from known data conditions; and (d) data stitching: stitch the data from x_{t-1} to x_{t+1} and x_{t-1} to x_t to obtain the predictions.

The deep characteristics extraction part utilizes a residual hybrid attention group (RHAG) to process intermediate characteristics and a final 3×3 convolution layer to aggregate information as Equation (10):

$$F_{DF} = H_{DF}(F_0) \quad (10)$$

where H_{DF} consists of RHAG and one 3×3 convolution layer.

The image reconstruction part combines shallow and deep characteristics and reconstructs the high-quality (HQ) image using a reconstruction module with pixel-shuffle for super-resolution tasks or two convolutions for same-resolution tasks as Equation (11):

$$I_{HQ} = H_{Rec}(F_0 + F_{DF}) \quad (11)$$

where $H_{Rec}(\cdot)$ denotes the reconstruction module.

The residual hybrid attention group (RHAG) comprises two main components: hybrid attention blocks (HAB) and overlapping cross-attention blocks (OCAB). The HAB module integrates self-attention and channel attention mechanisms, enabling the activation of a broader range of input pixels and boosting reconstruction performance. It employs a large window size for window-based self-attention and incorporates a channel attention block (CAB) to dynamically adjust feature scaling. On the other hand, the OCAB module improves cross-window interactions by applying overlapping windows to key and value features, ensuring that the query can access richer contextual information. Figure 4 shows transformer-based data restoration architecture specifically.

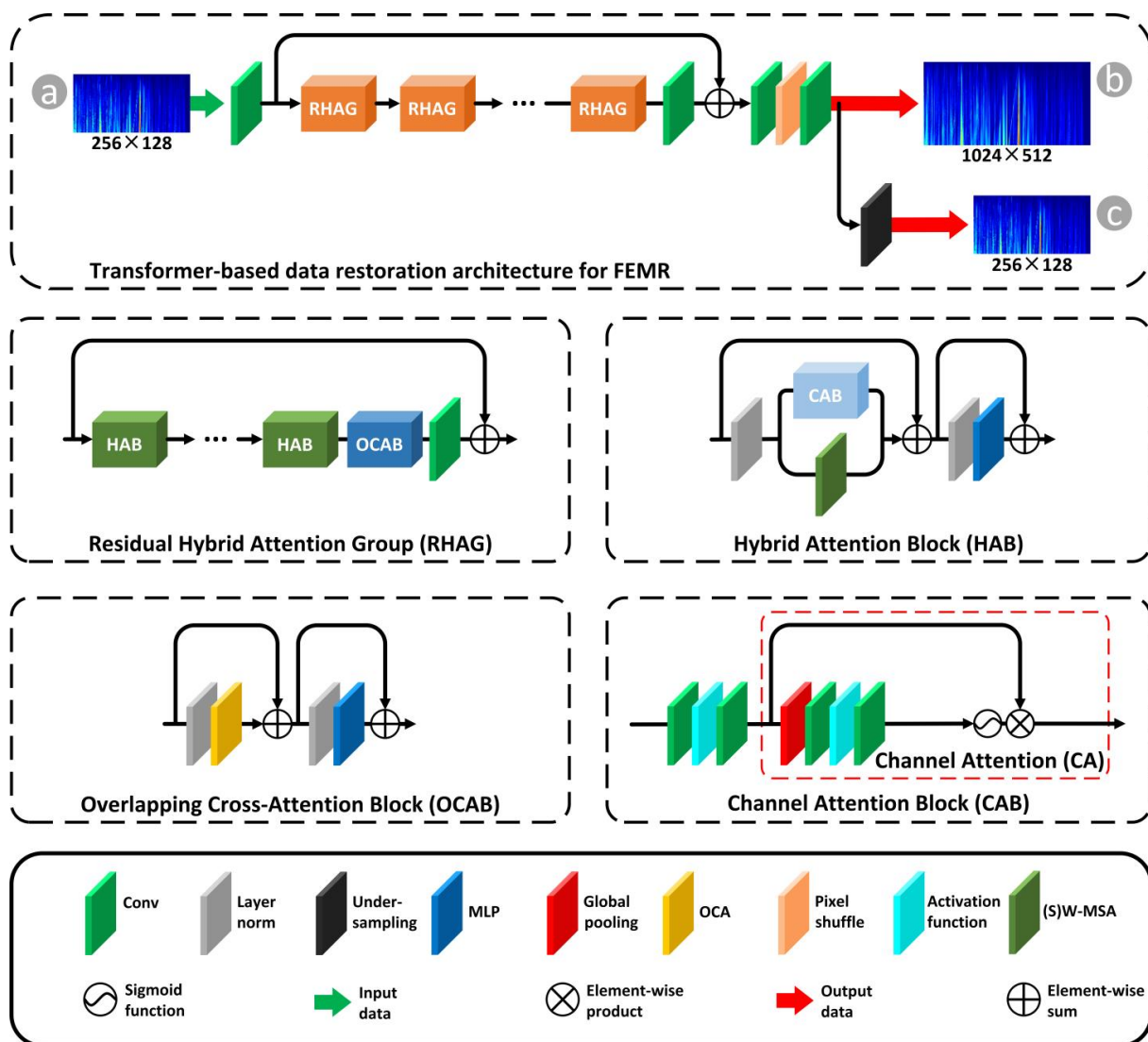


Figure 4. Schematic of transformer-based data restoration architecture. (a) Inputs: prediction of the guided diffusion model has the data size of 256×128 . (b) Outputs for next recurrent's prediction task: the trained data restoration architecture can reconstruct the time–frequency data lost during the prediction process. The size of the reconstructed time–frequency data is 1024×512 . In order to achieve continuous prediction using down-sampling, the pixel size is changed to 256×128 . (c) Outputs for precursor characteristics recognition task: to ensure the accuracy of precursor characteristics recognition, auxiliary model directly extracts characteristics from the reconstructed time–frequency data (1024×512) to achieve early warning of rock burst.

2.2.3. Recurrent Structure for Super Prediction

In this study, as shown in Figure 5, the super prediction approach refers to the use of recurrent structure to achieve multi-step prediction of precursor signals for rock outburst. This method not only predicts the signal for the next time step but also recursively generates predictions for multiple future time steps, thus providing a longer warning horizon.

The recursive super prediction strategy is a key innovation in this study. We split the time–frequency data into segments of specific durations. During model training, we use the previous segment of data as input to predict the next segment. Then, we use the prediction results from the previous step serves as input for the next prediction. Through this recursive process, the model generates predictions for multiple future time steps. By introducing the recurrent structure, this study achieved super prediction of FEMR signals, significantly improving both prediction time span.

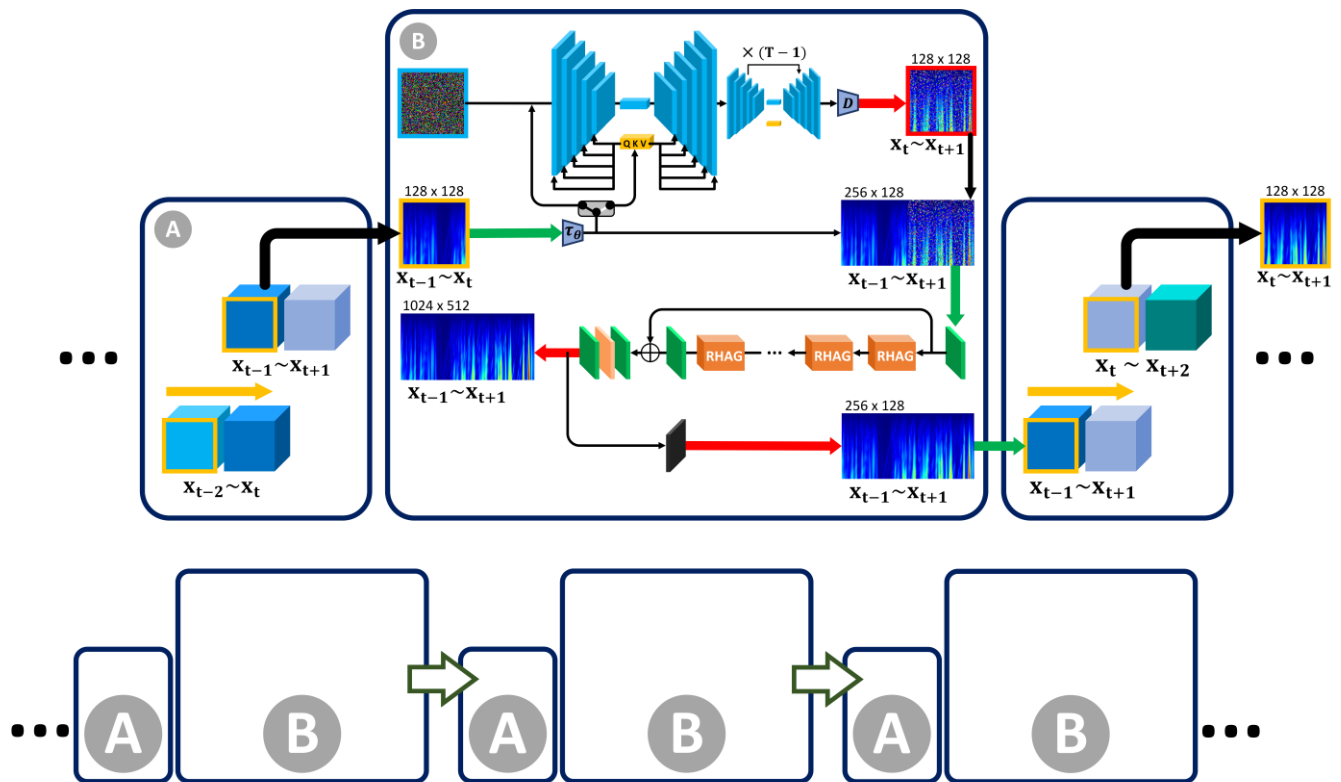


Figure 5. Schematic of recurrent structure: (A) window sliding operation: the data range of the window selection slides from $x_{t-2} \sim x_{t-1}$ ($x_{t-1} \sim x_t$) to $x_{t-1} \sim x_t$ ($x_t \sim x_{t+1}$)—the data range of the time-frequency data slides from $x_{t-2} \sim x_t$ ($x_{t-1} \sim x_{t+1}$) to $x_{t-1} \sim x_{t+1}$ ($x_t \sim x_{t+2}$); (B) add Gaussian noise to the ground truth time-frequency data that needs to be predicted; inputs: the known part of the time-frequency data and pure Gaussian noise; and network architecture of guided diffusion model with transformer.

2.3. Dataset Preparation

2.3.1. Acquisition of FEMR Data

As shown in Figure 6, the FEMR data from the NO.6, NO.7, and NO.8 return wind roadway in the West No. 3 mining area of the Buertai coal mine from January 2019 to May 2023 were selected as the research subject.

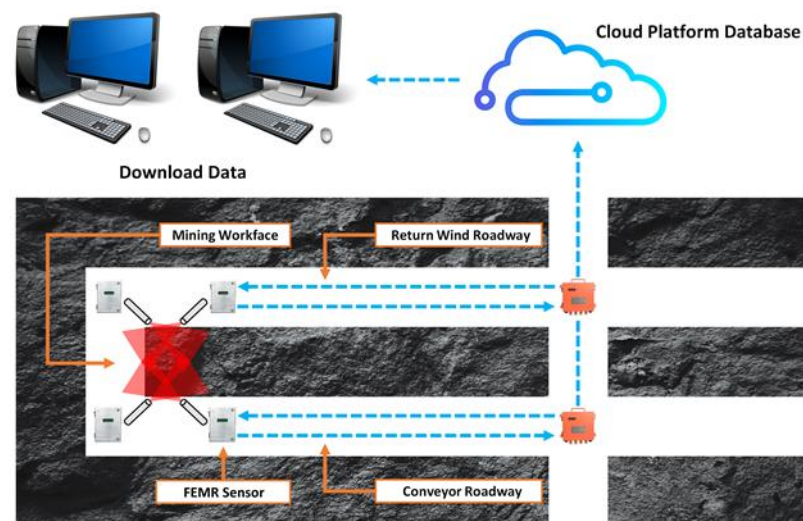


Figure 6. Schematic of recurrent structure.

2.3.2. Principle of Continuous Wavelet Transform (CWT)

Wavelet transforms have become essential tools in various signal-processing fields, with the continuous wavelet transform (CWT) standing out for its effectiveness in signal analysis. In CWT, the wavelet associated with scale a and time position b is expressed as Equation (12):

$$\varphi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \varphi\left(\frac{t-b}{a}\right) \quad (12)$$

where $\varphi(t)$ is the fundamental wavelet that acts as a 'prototype', serving as a bandpass filter. The factor $|a|^{-1/2}$ ensures energy preservation. The time-scale parameters (b, a) are discretized which can be approached in various ways, leading to different forms of wavelet transforms. Following the conventions of the classical Fourier transform, this research adopts a comparable framework. The concept of CWT was first introduced by Goupillaud. Both the time t and timescale parameters vary continuously as Equation (13):

$$CWT\{x(t); a, b\} = \int x(t) \varphi_{a,b}^*(t) dt \quad (13)$$

In Equation (13), the asterisk denotes the complex conjugate operation, and the wavelet sequence (WS) coefficients correspond to the sampled continuous wavelet transform (CWT) coefficients. Time is modeled as a continuous variable, whereas the timescale parameters (b, a) are defined within the timescale plane (b, a) .

To enhance data reliability and minimize errors or omissions during transmission, each time–frequency window contains 120 selectively sampled points. Furthermore, the implementation of overlapping data segments ensures seamless and continuous data analysis.

2.3.3. Time–Frequency Analysis of FEMR

The FEMR data were processed using CWT. Figure 7 shows the CWT data of FEMR signals with significant dynamic performance. The CWT analysis of FEMR data training samples in the database indicates that, in the abnormal state, the dominant amplitude band of the FEMR data is mainly concentrated between 80 to 230, as shown in Figure 7. The findings demonstrate that the time–frequency features of FEMR signals under normal conditions exhibit notable distinctions compared to those observed during abnormal conditions, enabling the detection of early warning signs preceding abrupt incidents.

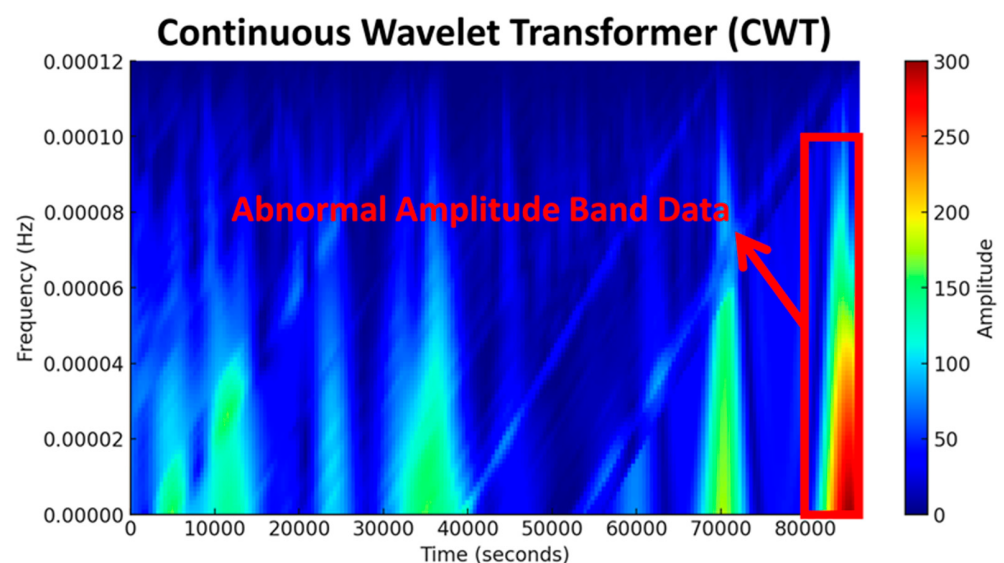


Figure 7. CWT time–frequency diagram of precursory characteristics.

Based on above analysis, we believe that the time–frequency characteristics of FEMR under normal conditions are significantly different from those in abnormal conditions, and can be used to recognize precursor characteristics of rock burst.

3. Results

3.1. Dataset Creation and Enhancement

The research focused on the Buertai coal mine in Erdos City, where uniform sensor models and standardized sampling procedures were implemented across all working faces. The investigation leveraged FEMR signal data spanning January 2019 through May 2023, extracted from the mine’s comprehensive database. Given the intermittent nature of rock burst precursor manifestations during extraction activities, the potential for model overfitting due to limited data samples was mitigated through strategic data expansion methodologies. These approaches were specifically designed to strengthen the predictive model’s resilience. To account for geographical variations in frequency distribution patterns that might affect precursor identification, the research team implemented advanced time–frequency characteristic enhancement strategies. These included the spatial transformation of graphical representations, selective image segmentation, and the incorporation of controlled Gaussian noise at varying intensities (0.05 and 0.005). Through these techniques, the repository of enhanced time–frequency diagrams was expanded to 2600 instances, substantially boosting the model’s detection accuracy.

The methodology incorporated a rigorous field data acquisition process coupled with meticulous data labeling, followed by the systematic partitioning of the dataset into distinct groups for model training, evaluation, and final assessment. Following established research conventions, the FEMR records from the Buertai mine were proportioned into training and validation cohorts at an 80:20 distribution. The comprehensive documentation of EMR signal classifications within the database, including specific numerical distributions for both training and validation groups, is presented in Table 1.

Table 1. The number of rock burst precursor training and validation sets in Buertai coal mine database.

Dataset	Buertai Coal Mine Database
	Rock Burst Precursor
Train	2080
Test	520

3.2. Signal Super Prediction and Precursor Recognition Framework Based on Guided Diffusion Model with Transformer

3.2.1. The Guided Diffusion Model

The guided diffusion model is primarily composed of a U-Net architecture, which includes an encoder, bottleneck, decoder, and output. The encoding component incorporates a feature extraction module based on convolutional operations with dimensions $3 \times 6 \times 64 \times 64$, residual modules with dimensions $3 \times 64 \times 64 \times 64$, $3 \times 128 \times 32 \times 32$, $3 \times 256 \times 16 \times 16$, and $3 \times 512 \times 8 \times 8$, down-sampling layers with dimensions $3 \times 64 \times 64 \times 64$, $3 \times 128 \times 32 \times 32$, and $3 \times 256 \times 16 \times 16$, and an attention module with dimension $3 \times 256 \times 16 \times 16$. The bottleneck includes residual modules with dimensions $3 \times 512 \times 8 \times 8$ and an attention module with dimension $3 \times 256 \times 8 \times 8$. The decoder comprises residual modules with dimensions $3 \times 1024 \times 8 \times 8$, $3 \times 384 \times 16 \times 16$, $3 \times 192 \times 32 \times 32$ and $3 \times 128 \times 64 \times 64$, as well as up-sampling layers with dimensions $3 \times 256 \times 8 \times 8$, $3 \times 128 \times 16 \times 16$, and $3 \times 64 \times 32 \times 32$. Finally, the output module is a convolutional layer with dimension $3 \times 64 \times 64 \times 64$. The prediction steps of the guided diffusion model in the training set are shown in Figure 8.

The guided diffusion model's parameter optimization was accomplished through the implementation of stochastic gradient descent (SGD) as the optimization algorithm. The training configuration employed a batch size of 8, while maintaining a dropout probability of 0.2 to mitigate the risk of overfitting. The learning rate was specifically tuned to 5×10^{-5} for optimal convergence.

Throughout the training process, testing was performed every 10 epochs. Due to disk storage limitations, training was halted after 4 days and 13 h, reaching epoch 2428. The LOSS curve is a crucial tool for understanding the model's training progress. Typically, during the early stages of training, the LOSS value is relatively high, as the model's parameters are far from optimal. As training progresses, the model begins to adjust its weights and biases, leading to a decrease in the LOSS value. This downward trend reflects the model's improvement in predicting the target outputs. Following a thorough evaluation of the loss values, the optimal model was saved at epoch 1850. The evaluation metrics, calculated through mean squared error (MSE) methodology for both training and validation datasets, demonstrated a consistent downward trajectory before reaching a plateau. This observed behavior suggests a successful convergence and the optimal adaptation of the Guided Diffusion architecture. The corresponding training loss progression is visually represented in Figure 9. All training procedures were executed on an RTX 5000 GPU equipped with 16 GB of VRAM (Nvidia, Santa Clara, CA, USA).

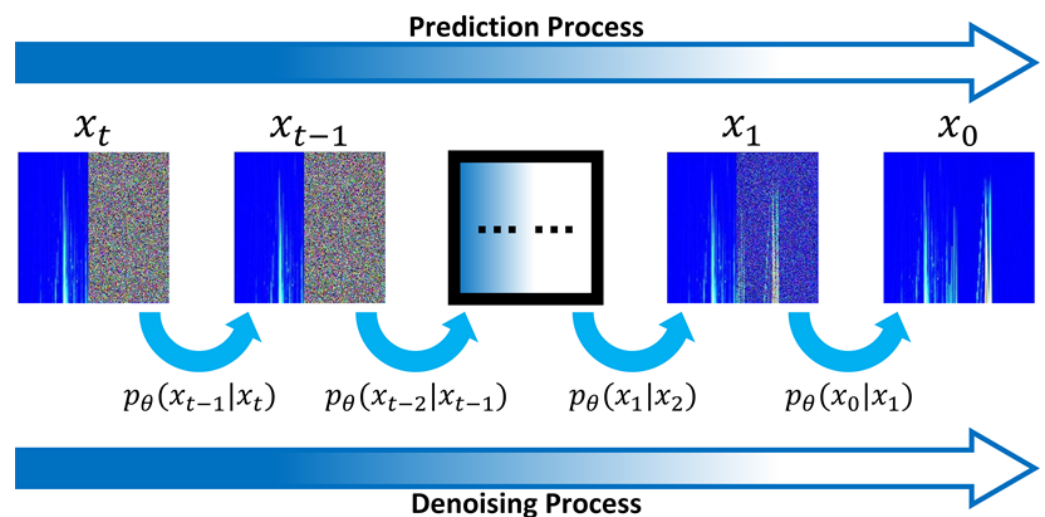


Figure 8. Diffusion model prediction process during training.

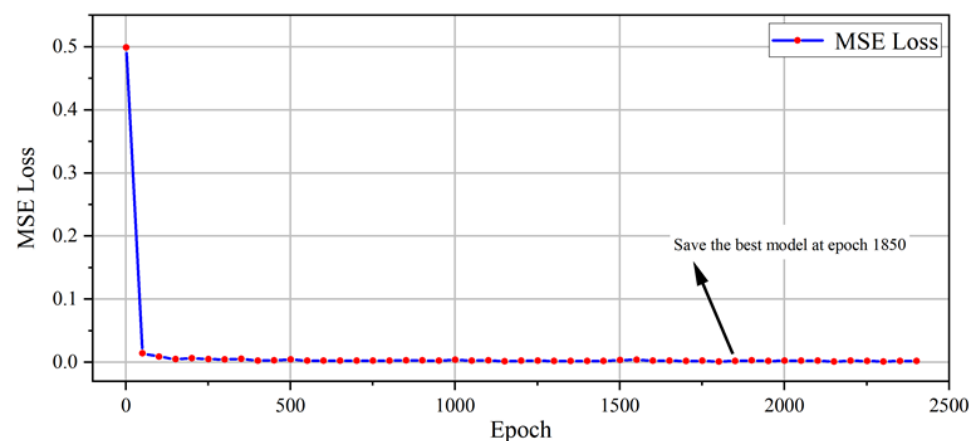


Figure 9. The loss curve for the training set.

3.2.2. Transformer-Based Data Restoration and Recurrent Structure for Super Prediction

We use a transformer-based data restoration architecture to reconstruct the FEMR data predicted by the guided diffusion model. The recurrent structure can predict data for 360 h (15 days) in the future based on 12 h of FEMR data, due to the potential FEMR data missing during the recurrent prediction process. As shown in Figure 5, the Transformer-based data reconstruction architecture reads FEMR data with dimensions $256 \times 128 \times 1$, and the output FEMR data with dimensions $256 \times 128 \times 1$ and $1024 \times 512 \times 1$. FEMR data with dimensions $256 \times 128 \times 1$ are used for the next restoration, while FEMR data with dimensions $1024 \times 512 \times 1$ are used for rock burst precursor recognition. Based on the past 12 h of FEMR data, we achieved predictions for the next 360 h (15 days).

However, during the fourth prediction reconstruct, the data become challenging to restore, making it difficult to recognize precursor characteristics of rock burst. If the known data are compressed along the temporal dimension, it is theoretically possible to achieve predictions for a longer future time frame. All testing processes were completed on an RTX 5000 GPU with 16 GB of VRAM.

3.2.3. Auxiliary Model for Recognizing Rock Burst Precursor Characteristics

The input FEMR data for the auxiliary model has dimensions of $1024 \times 512 \times 3(H \times W \times C)$. Initially, the data are processed using a 4×4 non-overlapping convolution with a stride of 4, followed by a normalization layer to generate $\frac{H}{4} \times \frac{W}{4} \times C$ patch embeddings. The architectural framework for building a multi-level feature hierarchy incorporates four distinct processing stages within its core structure, with each successive stage implementing incrementally larger stride values. Between consecutive stages, a non-overlapping 2×2 convolution with a stride of 2 is applied to down-sample the characteristics map, reducing its spatial size by half and doubling the characteristics dimensions. For the classification task, the characteristics maps from the final stage are normalized, and a linear classifier with pooled characteristics is used to predict the logits.

In Stage 1, the number of blocks is 1, the channel dimension is 96, the region size in the local attention module is 7, and the number of heads in DMHA is 3. In Stage 2, the number of blocks is 1, the channel dimension is 192, the region size in the local attention module is 7, and the number of heads in DMHA is 6. In Stage 3, the number of blocks is 3, the channel dimension is 384, the region size in the local attention module is 7, the number of heads in DMHA is 12, and the offset groups in DMHA is 3. In Stage 4, the number of blocks is 1, the channel dimension is 768, the region size in the local attention module is 7, the number of heads in DMHA is 24, and the offset groups in DMHA is 6.

An auxiliary model, which has a depth of 225 layers and approximately 11×10^6 parameters, was used. The model's parameter optimization was accomplished through the implementation of the stochastic gradient descent algorithm. The model was configured with a batch size of 3, a learning rate set at 1×10^{-2} , and a dropout rate of 0.9 to prevent overfitting. Throughout the training process, testing was conducted every epoch. Due to disk capacity constraints, the training was stopped after 55.2 min, at epoch 100. After the comprehensive verification of the loss value, the best training model was saved at epoch 57. The training metrics, calculated through binary cross-entropy (BCE) for both the training and validation datasets, displayed a consistent reduction before reaching equilibrium. This behavior confirms the successful convergence and optimal adaptation of the guided diffusion architecture. The corresponding training loss progression is illustrated in Figure 10. All training and testing processes were completed on an RTX 5000 GPU with 16 GB of VRAM. During the testing process, 510 out of 520 FEMR data samples were correctly recognized, resulting in an accuracy of 98.07%.

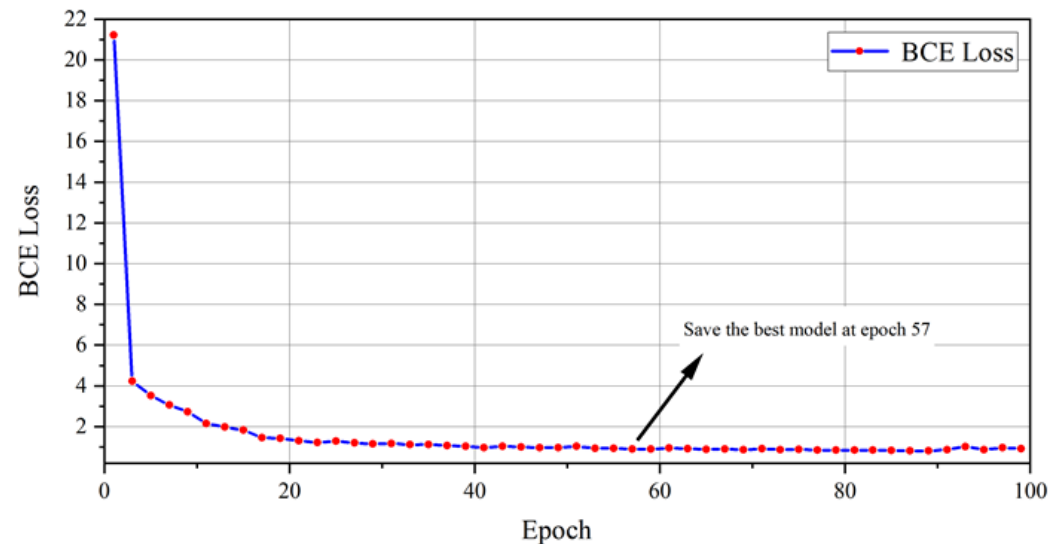


Figure 10. The loss curve for the training set.

3.3. Early Warning Results for Rock Burst Using Signal Super Prediction and Precursor Recognition Framework

We selected 15 days of continuous FEMR data with rock burst risk, recorded from 1 April to 15 April 2023, from the No. 8 return wind roadway in the West No. 3 mining area of the Buertai coal mine, to test the framework. FEMR time-series data span from 00:00 on 1 April 2023, to 22:00 on 7 April 2023, during which the FEMR time-series data exhibit a periodic cyclic trend, indicating active mining operations at the working face. From 00:00 on 1 April 2023, to 15:51 on 9 April 2023, a slight cyclic growth trend in the FEMR time-series data can be observed. However, by 06:13 on 12 April 2023, the FEMR time-series data show a significant and pronounced cyclic growth trend, which can be identified as a distinctive signature preceding rock burst occurrences within the FEMR signal patterns.

We input the known FEMR data from 00:00 to 12:00 on 1 April 2023, into the super prediction framework to predict future data. The FEMR time-series data from April 1 to 15 April 2023 is shown in Figure 11, while the time-frequency data for the same period are presented in Figure 11. The results of the signal super prediction are shown in Figure 12, where we successfully predicted 360 h (15 days) of future FEMR data based on 12 h of known data. The prediction results demonstrate that the framework can achieve long-term data predicting using a relatively small amount of initial FEMR time-series data. We input the FEMR time-frequency data obtained from the framework's super prediction results into an auxiliary model, which successfully recognized precursor characteristics of rock burst, and the confidence level for recognizing precursor characteristics is 0.66. The framework's super prediction and precursor characteristics recognition results are shown in Figure 12. It can be observed that the precursor characteristics of FEMR exhibit a relatively high amplitude.

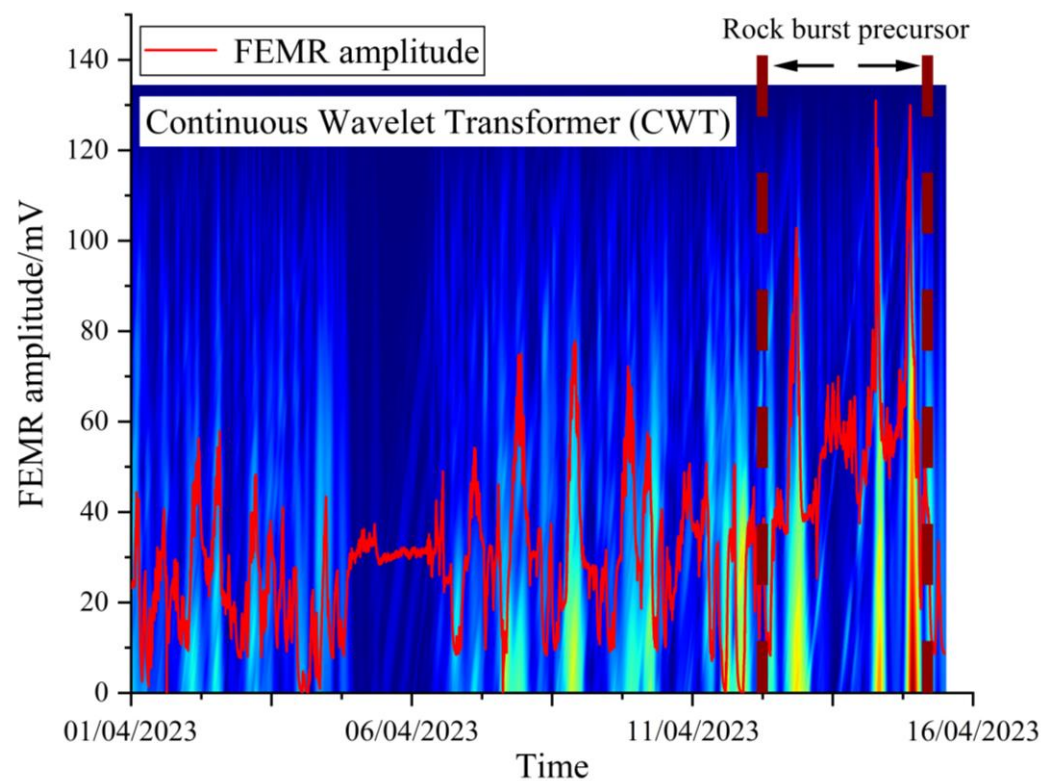


Figure 11. Comparison chart of raw data and CWT time–frequency data.

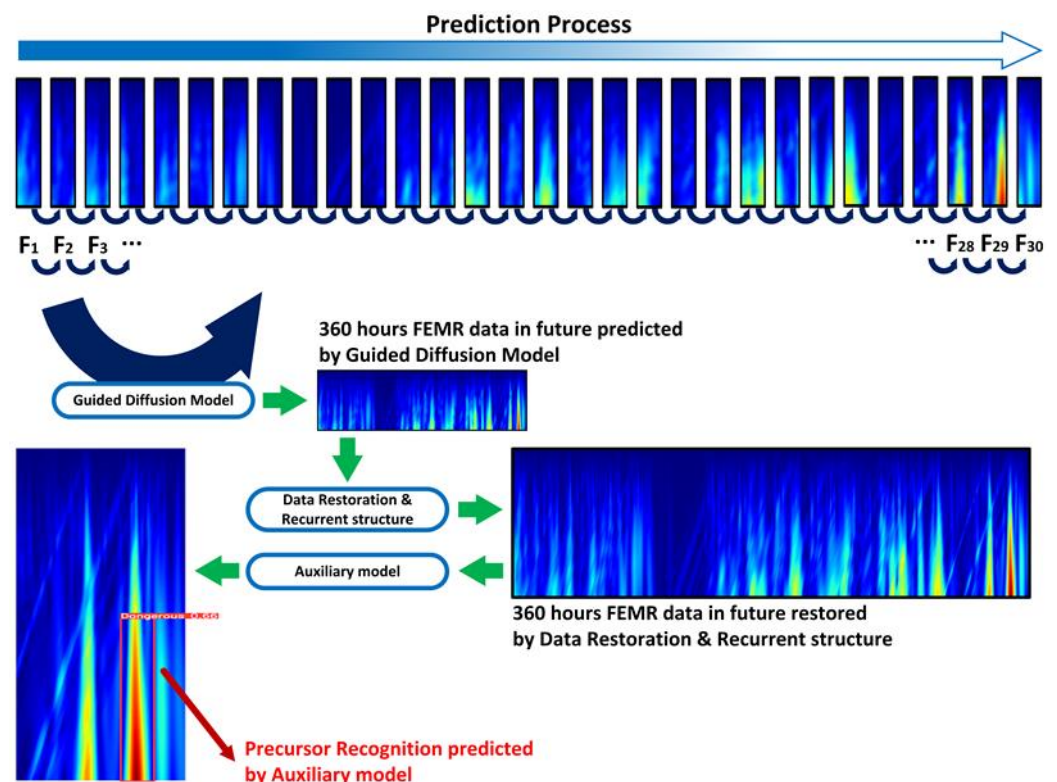


Figure 12. The framework's super prediction and precursor characteristics recognition results.

4. Conclusions

The traditional method of early warning through precursor recognition typically offers limited lead time, usually within 7 days. Prolonged precursor signals can delay warnings, posing challenges for timely intervention. Some researchers have utilized recurrent neural

networks (RNNs) and transformer-based architectures for time-series prediction, achieving FEMR signal forecasting up to 7 days in advance. While effective for short-term predictions, this approach necessitates multi-source data fusion and extensive sensor deployments to ensure a high accuracy, relying on substantial pre-existing datasets. This requirement imposes significant storage burdens on coal mine databases, a persistent challenge in operational contexts.

In contrast, this study introduces an innovative approach leveraging a guided diffusion model coupled with a transformer-based data restoration architecture for the processing of known FEMR signals. This model facilitates the super prediction of FEMR data, forecasting up to 360 h (15 days) into the future based solely on 12 h of known data. Additionally, an auxiliary model is employed to identify rock burst precursor characteristics within the FEMR data. The empirical results underscore the efficacy of this approach in predicting FEMR signals over extended time periods, while also conducting a comprehensive analysis of the upward trends inherent in FEMR signals—key indicators of rock burst precursors.

The academic contribution of this method lies in its ability to extend prediction horizons and enhance the interpretability of rock burst precursor identification. With an accuracy rate of 98.07%, the approach significantly improves the reliability of early warning systems, advancing both the forecasting capability and risk assessment of rock burst hazards up to 15 days ahead. In conclusion, this study offers a novel and rigorous methodology that not only pushes the boundaries of prediction timelines but also contributes to the advancement of safety management and monitoring systems in coal mining operations, enhancing the accuracy and reliability of rock burst early warnings.

Author Contributions: M.W.: conceptualization, methodology, and writing—review and editing. Z.D.: methodology, and writing—original draft. C.C.: methodology, resources, and supervision. E.W.: methodology and validation. H.J.: methodology and validation. X.L.: methodology and validation. J.W.: methodology and validation. G.S.: methodology and validation. Y.L.: methodology and validation. All authors have read and agreed to the published version of the manuscript.

Funding: The National Natural Science Foundation of China (52227901 and 51934007); and The National Key Research and Development Program of China (2022YFC3004705).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The research data underlying the conclusions of this investigation are accessible through the Buertai Coal Mine Data Cloud Platform. Access to these datasets is subject to specific restrictions, as they were utilized under authorized licensing agreements for the current research. Interested parties may request access to the data from the corresponding authors, pending approval from the Buertai Coal Mine Data Cloud Platform administration.

Acknowledgments: The authors gratefully acknowledge the financial support of the above-mentioned agencies. In addition, the authors would like to express their special thanks to the journal editors and reviewers for their hard work.

Conflicts of Interest: Authors Mingyue Weng, Chuncheng Cai and Yong Liu were employed by the company Shanghai Datun Energy Resources Co., Ltd. Author Guorui Su was employed by the Information Institution of Ministry of Emergency Management. 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.

References

1. Zhang, G.; Wang, E. Risk identification for coal and gas outburst in underground coal mines: A critical review and future directions. *Gas Sci. Eng.* **2023**, *118*, 205106. [\[CrossRef\]](#)
2. Gu, S.; Zhang, W.; Jiang, B.; Hu, C. Case of rock burst danger and its prediction and prevention in tunneling and mining period at an irregular coal face. *Geotech. Geol. Eng.* **2019**, *37*, 2545–2564. [\[CrossRef\]](#)
3. Yin, S.; Wang, E.; Li, Z.; Zang, Z.; Liu, X.; Zhang, C.; Ding, X.; Aihemaiti, A. Multifractal and b-value nonlinear time-varying characteristics of acoustic emission for coal with different impact tendency. *Measurement* **2025**, *248*, 116896. [\[CrossRef\]](#)
4. Li, X.; Wang, E.; Li, Z.; Liu, Z.; Song, D.; Qiu, L. Rock burst monitoring by integrated microseismic and electromagnetic radiation methods. *Rock Mech. Rock Eng.* **2016**, *49*, 4393–4406. [\[CrossRef\]](#)
5. Di, Y.; Wang, E. Rock burst precursor electromagnetic radiation signal recognition method and early warning application based on recurrent neural networks. *Rock Mech. Rock Eng.* **2021**, *54*, 1449–1461. [\[CrossRef\]](#)
6. Das, S.; Frid, V.; Rabinovitch, A.; Bahat, D.; Kushnir, U. Insights into the Dead Sea Transform Activity through the study of fracture-induced electromagnetic radiation (FEMR) signals before the Syrian-Turkey earthquake (Mw-6.3) on 20.2.2023. *Sci. Rep.* **2024**, *14*, 4579. [\[CrossRef\]](#)
7. Bai, W.; Shu, L.; Sun, R.; Xu, J.; Silberschmidt, V.V.; Sugita, N. Mechanism of material removal in orthogonal cutting of cortical bone. *J. Mech. Behav. Biomed. Mater.* **2020**, *104*, 103618. [\[CrossRef\]](#)
8. Frid, V.; Rabinovitch, A.; Bahat, D.; Kushnir, U. Fracture electromagnetic radiation induced by a seismic active zone (in the Vicinity of Eilat City, Southern Israel). *Remote Sens.* **2023**, *15*, 3639. [\[CrossRef\]](#)
9. Wang, X.; Yu, S.; Yin, D. Time-frequency Characteristics of Micro-seismic Signals Before and after Rock Burst. In Proceedings of the 9th China-Russia Symposium “Coal in the 21st Century: Mining, Intelligent Equipment and Environment Protection” (COAL 2018), Qingdao, China, 18–21 October 2018; Atlantis Press: Amsterdam, The Netherlands, 2018; pp. 9–12.
10. Ma, T.; Lin, D.; Tang, L.; Li, L.; Tang, C.; Yadav, K.P.; Jin, W. Characteristics of rockburst and early warning of microseismic monitoring at qinling water tunnel. *Geomat. Nat. Hazards Risk* **2022**, *13*, 1366–1394. [\[CrossRef\]](#)
11. Zhang, Z.; Ye, Y.; Luo, B.; Chen, G.; Wu, M. Investigation of microseismic signal denoising using an improved wavelet adaptive thresholding method. *Sci. Rep.* **2022**, *12*, 22186. [\[CrossRef\]](#)
12. Wu, M.; Ye, Y.; Wang, Q.; Hu, N. Development of rockburst research: A comprehensive review. *Appl. Sci.* **2022**, *12*, 974. [\[CrossRef\]](#)
13. Zhang, S.; Tang, C.; Wang, Y.; Li, J.; Ma, T.; Wang, K. Review on early warning methods for rockbursts in tunnel engineering based on microseismic monitoring. *Appl. Sci.* **2021**, *11*, 10965. [\[CrossRef\]](#)
14. Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E. Deep learning for computer vision: A brief review. *Comput. Intell. Neurosci.* **2018**, *2018*, 7068349. [\[CrossRef\]](#)
15. Hewamalage, H.; Bergmeir, C.; Bandara, K. Recurrent neural networks for time series forecasting: Current status and future directions. *Int. J. Forecast.* **2021**, *37*, 388–427. [\[CrossRef\]](#)
16. Ahmed, S.; Nielsen, I.E.; Tripathi, A.; Siddiqui, S.; Ramachandran, R.P.; Rasool, G. Transformers in time-series analysis: A tutorial. *Circuits Syst. Signal Process.* **2023**, *42*, 7433–7466. [\[CrossRef\]](#)
17. Wang, Z.; She, Q.; Ward, T.E. Generative adversarial networks in computer vision: A survey and taxonomy. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–38. [\[CrossRef\]](#)
18. Du, Z.; Liu, X.; Wang, J.; Jiang, G.; Meng, Z.; Jia, H.; Xie, H.; Zhou, X. Response characteristics of gas concentration level in mining process and intelligent recognition method based on bi-lstm. *Min. Metall. Explor.* **2023**, *40*, 807–818. [\[CrossRef\]](#)
19. Zhang, S.; Mu, C.; Feng, X.; Ma, K.; Guo, X.; Zhang, X. Intelligent dynamic warning method of rockburst risk and level based on recurrent neural network. *Rock Mech. Rock Eng.* **2024**, *57*, 3509–3529. [\[CrossRef\]](#)
20. Wojtecki, Ł.; Iwaszenko, S.; Apel, D.B.; Cichy, T. An attempt to use machine learning algorithms to estimate the rockburst hazard in underground excavations of hard coal mine. *Energies* **2021**, *14*, 6928. [\[CrossRef\]](#)
21. Mahmood, M.R.; Matin, M.A.; Sarigiannidis, P.; Goudos, S.K. A comprehensive review on artificial intelligence/machine learning algorithms for empowering the future IoT toward 6G era. *IEEE Access* **2022**, *10*, 87535–87562. [\[CrossRef\]](#)
22. Basnet, P.M.S.; Mahtab, S.; Jin, A. A comprehensive review of intelligent machine learning based predicting methods in long-term and short-term rock burst prediction. *Tunn. Undergr. Space Technol.* **2023**, *142*, 105434. [\[CrossRef\]](#)
23. Prasad, S.C.; Prasad, P. Deep recurrent neural networks for time series prediction. *arXiv* **2014**, arXiv:1407.5949.
24. Huang, L.; Xiang, L. Method for meteorological early warning of precipitation-induced landslides based on deep neural network. *Neural Process. Lett.* **2018**, *48*, 1243–1260. [\[CrossRef\]](#)
25. Deb, S.; Chanda, A.K. Comparative analysis of contextual and context-free embeddings in disaster prediction from Twitter data. *Mach. Learn. Appl.* **2022**, *7*, 100253. [\[CrossRef\]](#)
26. Jiao, H.; Song, W.; Cao, P.; Jiao, D. Prediction method of coal mine gas occurrence law based on multi-source data fusion. *Heliyon* **2023**, *9*, e17117. [\[CrossRef\]](#)
27. Kiziroglou, M.E.; Boyle, D.E.; Yeatman, E.M.; Cilliers, J.J. Opportunities for sensing systems in mining. *IEEE Trans. Ind. Inform.* **2016**, *13*, 278–286. [\[CrossRef\]](#)

28. Liu, M.Y.; Tuzel, O. Coupled generative adversarial networks. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Newry, UK, 2016; Volume 29.
29. Adler, J.; Lunz, S. Banach wasserstein gan. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Newry, UK, 2018; Volume 31.
30. Brock, A. Large Scale GAN Training for High Fidelity Natural Image Synthesis. *arXiv* **2018**, arXiv:1809.11096.
31. Zhang, H.; Goodfellow, I.; Metaxas, D.; Odena, A. Self-attention generative adversarial networks. In *Proceedings of the International Conference on Machine Learning*, PMLR, 2019, Long Beach, CA, USA, 9–15 June 2019; pp. 7354–7363.
32. Mirza, M. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.
33. Salimans, T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; Chen, X.; Chen, X. Improved techniques for training gans. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Newry, UK, 2016; Volume 29.
34. Radford, A. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434.
35. Nagpal, S.; Verma, S.; Gupta, S.; Aggarwal, S. A guided learning approach for generative adversarial networks. In *Proceedings of the 2020 IEEE International Joint Conference on Neural Networks (IJCNN)*, Glasgow, UK, 19–24 July 2020; pp. 1–8.
36. Karras, T. A Style-Based Generator Architecture for Generative Adversarial Networks. *arXiv* **2019**, arXiv:1812.04948.
37. Younesi, A.; Ansari, M.; Fazli, M.; Ejlali, A.; Shafique, M.; Henkel, J. A Comprehensive Survey of Convolutions in Deep Learning: Applications, Challenges, and Future Trends. *IEEE Access* **2024**, *12*, 41180–41218. [[CrossRef](#)]
38. Panariello, M. *Low-Complexity Neural Networks for Robust Acoustic Scene Classification in Wearable Audio Devices*; Politecnico di Torino: Turin, Italy, 2022.
39. Peng, J.; Liu, Y.; Wang, M.; Li, Y.; Li, H. Zero-Shot Self-Consistency Learning for Seismic Irregular Spatial Sampling Reconstruction. *arXiv* **2024**, arXiv:2411.00911.
40. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Newry, UK, 2014; Volume 27.
41. Cao, H.; Tan, C.; Gao, Z.; Xu, Y.; Chen, G.; Heng, P.-A.; Li, S.Z. A survey on generative diffusion models. *IEEE Trans. Knowl. Data Eng.* **2024**, *36*, 2814–2830. [[CrossRef](#)]
42. Yang, L.; Zhang, Z.; Song, Y.; Hong, S.; Xu, R.; Zhao, Y.; Zhang, W.; Cui, B.; Yang, M.-H. Diffusion models: A comprehensive survey of methods and applications. *ACM Comput. Surv.* **2023**, *56*, 1–39. [[CrossRef](#)]
43. Dhariwal, P.; Nichol, A. Diffusion models beat gans on image synthesis. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Newry, UK, 2021; Volume 34, pp. 8780–8794.
44. Srivastava, A.; Sutton, C. Autoencoding variational inference for topic models. *arXiv* **2017**, arXiv:1703.01488.
45. Ramesh, A.; Pavlov, M.; Goh, G.; Gray, S.; Voss, C.; Radford, A.; Chen, M.; Sutskever, I. Zero-shot text-to-image generation. In *Proceedings of the International Conference on Machine Learning*, PMLR, Virtual, 18–24 July 2021; pp. 8821–8831.
46. Zhu, L.; Zheng, Y.; Zhang, Y.; Wang, X.; Wang, L.; Huang, H. Temporal Residual Guided Diffusion Framework for Event-Driven Video Reconstruction. In *Proceedings of the European Conference on Computer Vision*; Springer: Cham, Switzerland, 2025; pp. 411–427.
47. Huang, Y.; Huang, J.; Liu, Y.; Yan, M.; Lv, J.; Liu, J.; Xiong, W.; Zhang, H.; Chen, S.; Cao, L. Diffusion model-based image editing: A survey. *arXiv* **2024**, arXiv:2402.17525. [[CrossRef](#)] [[PubMed](#)]
48. Fuest, M.; Ma, P.; Gui, M.; Schusterbauer, J.; Hu, V.T.; Ommer, B. Diffusion models and representation learning: A survey. *arXiv* **2024**, arXiv:2407.00783.
49. Blattmann, A.; Rombach, R.; Ling, H.; Dockhorn, T.; Kim, S.W.; Fidler, S.; Kreis, K. Align your latents: High-resolution video synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver, BC, Canada, 18–22 June 2023; pp. 22563–22575.
50. Saharia, C.; Chan, W.; Chang, H.; Lee, C.; Ho, J.; Salimans, T.; Fleet, D.; Norouzi, M. Palette: Image-to-image diffusion models. In *Proceedings of the ACM SIGGRAPH 2022 Conference Proceedings*, Vancouver, BC, Canada, 7–11 August 2022; pp. 1–10.
51. Yu, J.; Lin, Z.; Yang, J.; Shen, X.; Lu, X.; Huang, T.S. Generative image inpainting with contextual attention. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5505–5514.
52. Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, LA, USA, 19–20 June 2022; pp. 10684–10695.
53. Avrahami, O.; Fried, O.; Lischinski, D. Blended latent diffusion. *ACM Trans. Graph. (TOG)* **2023**, *42*, 1–11. [[CrossRef](#)]
54. Lovelace, J.; Kishore, V.; Wan, C.; Shekhtman, E.; Weinberger, K.Q. Latent diffusion for language generation. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Newry, UK, 2024; Volume 36.
55. Xu, M.; Powers, A.S.; Dror, R.O.; Ermon, S.; Leskovec, J. Geometric latent diffusion models for 3d molecule generation. In *Proceedings of the International Conference on Machine Learning*. PMLR, Honolulu, HI, USA, 23–29 July 2023; pp. 38592–38610.

56. Song, Y.; Ermon, S. Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Newry, UK, 2019; Volume 32.
57. Nichol, A.Q.; Dhariwal, P. Improved denoising diffusion probabilistic models. In Proceedings of the International Conference on Machine Learning, PMLR, Virtual, 18–24 July 2021; pp. 8162–8171.
58. Sohl-Dickstein, J.; Weiss, E.; Maheswaranathan, N.; Ganguli, S. Deep unsupervised learning using nonequilibrium thermodynamics. In Proceedings of the International Conference on Machine Learning, PMLR, Lille, France, 6–11 July 2015; pp. 2256–2265.
59. Kingma, D.P. Auto-encoding variational bayes. *arXiv* **2013**, arXiv:1312.6114.
60. Ahmed, D.M.; Hassan, M.M.; Mstafa, R.J. A review on deep sequential models for forecasting time series data. *Appl. Comput. Intell. Soft Comput.* **2022**, 2022, 6596397. [\[CrossRef\]](#)
61. Mavrogiorgos, K.; Kiourtis, A.; Mavrogiorgou, A.; Gucek, A.; Menychtas, A.; Kyriazis, D. Mitigating Bias in Time Series Forecasting for Efficient Wastewater Management. In Proceedings of the 2024 7th International Conference on Informatics and Computational Sciences (ICICoS), Semarang, Indonesia, 17–18 July 2024; IEEE: New York, NY, USA, 2024; pp. 185–190.
62. Nguyen, H.P.; Liu, J.; Zio, E. A long-term prediction approach based on long short-term memory neural networks with automatic parameter optimization by Tree-structured Parzen Estimator and applied to time-series data of NPP steam generators. *Appl. Soft Comput.* **2020**, 89, 106116. [\[CrossRef\]](#)
63. Rahman, M.; Shakeri, M.; Khatun, F.; Tiong, S.K.; Alkahtani, A.A.; Samsudin, N.A.; Amin, N.; Pasupuleti, J.; Hasan, M.K. A comprehensive study and performance analysis of deep neural network-based approaches in wind time-series forecasting. *J. Reliab. Intell. Environ.* **2023**, 9, 183–200. [\[CrossRef\]](#)
64. Vaswani, A. Attention is all you need. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Newry, UK, 2017.
65. Han, K.; Wang, Y.; Chen, H.; Chen, X.; Guo, J.; Liu, Z. A survey on vision transformer. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, 45, 87–110. [\[CrossRef\]](#) [\[PubMed\]](#)
66. Parmar, N.; Vaswani, A.; Uszkoreit, J.; Kaiser, L.; Shazeer, N.; Ku, A.; Tran, D. Image transformer. In Proceedings of the International Conference on Machine Learning, PMLR, Stockholm, Sweden, 10–15 July 2018; pp. 4055–4064.
67. Liu, Z.; Ning, J.; Cao, Y.; Wei, Y.; Zhang, Z.; Lin, S.; Hu, H. Video swin transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–20 June 2022; pp. 3202–3211.
68. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 10012–10022.
69. Liu, Z.; Hu, H.; Lin, Y.; Yao, Z.; Xie, Z.; Wei, Y.; Ning, J.; Cao, Y.; Zhang, Z.; Dong, L.; et al. Swin transformer v2: Scaling up capacity and resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–20 June 2022; pp. 12009–12019.
70. Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; Timofte, R. Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 1833–1844.
71. Ho, J.; Jain, A.; Abbeel, P. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Newry, UK, 2020; Volume 33, pp. 6840–6851.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.