electronics

MDPI

Article

Hybrid Deep Learning Method for Vibration-Based Gear Fault Diagnosis in Shearer Rocker Arm

Joshua Fenuku [1], Hua Ding [1,*], Gertrude Selase Gosu [2], Xiaochun Sun [1] and Ning Li [1]

[1] College of Mechanical and Vehicle Engineering, Taiyuan University of Technology, Taiyuan 030024, China; zhoushuhua0035@link.tyut.edu.cn (J.F.); sunxiaochun0022@link.tyut.edu.cn (X.S.); lining0013@link.tyut.edu.cn (N.L.)

[2] College of Computer Science, Taiyuan University of Technology, Taiyuan 030024, China; gaoshuting0007@link.tyut.edu.cn

* Correspondence: dinghua2002@163.com

Abstract

In underground coal mining, the gear of a shearer's rocker arm endures extreme stress and environmental fluctuations. Failures in this vital component can pose serious safety hazards, cause prolonged operational downtime, and result in significant financial losses. Therefore, accurate gear fault diagnosis is crucial. However, conventional diagnostic methods often struggle with limited feature extraction and poor performance when dealing with non-stationary, noisy signals typical of this environment. To address these challenges, a hybrid model consisting of Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM) network, and Markov Transition Model (MTM) is proposed. In this framework, the CNN is used to extract both global and local features related to gear fault. A time-distributed feature extractor is then integrated with the LSTM to capture the temporal progression of these features, aiding in effective modeling of fault evolution over time. Finally, the MTM further refines classification by incorporating probabilistic state transition between fault conditions, thereby improving diagnostic stability and robustness under noise. Experimental validation was done using vibration data from the Taizhong Coal Machinery rocker arm test platform and gear data from Southeast University and achieved up to 99.79% accuracy. These results show this proposed method outperformed other advanced diagnostic methods, offering dependable fault diagnosis and strong noise resistance even under extreme noise conditions of −5 dB SNR.

Keywords: fault diagnosis; Convolutional Neural Network; Long Short-Term Memory; Markov Transition Model; shearer rocker arm; deep learning; vibration measurement

Check for updates

Academic Editor: Ahmed Abu-Siada

Received: 15 March 2026
Revised: 2 April 2026
Accepted: 4 April 2026
Published: 10 April 2026

Copyright: © 2026 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.

1. Introduction

Shearers are essential equipment in modern coal mining, enduring tough, ever-changing conditions that accelerate wear and damage of drivetrain parts [1]. The gears inside the rocker arm are especially vulnerable to failure due to high loads, exposure to abrasive coal dust, severe shock impacts, and rapid shifts in operational requirements [2]. These issues lead to early fatigue and breakdown, highlighting the urgent need for precise and early gear fault diagnosis to protect miners, prevent major equipment failures, and reduce costly downtime [3]. The complex nature of shearer systems and frequent component failures emphasize the need for advanced fault detection methods to ensure smooth operation and avoid costly financial losses [4]. Therefore, creating a trustworthy, intelligent fault diagnosis system for shearer rocker arm gears is crucial, especially given the high

background noise and tough conditions underground, which make robust and efficient diagnostics even more important [5].

Traditional fault diagnosis methods for rotating machinery typically depend on signal processing techniques like Fast Fourier Transform (FFT), wavelet transforms, and empirical mode decomposition (EMD) to extract fault-specific features from vibration or acoustic signals [1]. Although these methods work well in controlled settings, they usually need extensive domain expertise for feature selection and are sensitive to the non-stationary and nonlinear traits of signals from machinery operating under varying loads, common in coal mining [6]. The high background noise further complicates feature extraction, often hiding subtle signs of early faults. With the rise in artificial intelligence, data-driven approaches have become more popular. Machine learning models such as Support Vector Machines (SVM) and Artificial Neural Networks (ANN) have been used for fault diagnosis, but they often depend on manually crafted features, which makes the process labor-intensive and less effective [7]. Deep learning, a branch of machine learning, shows great potential in fault detection by effectively handling complex data and automatically learning layered feature representations from raw signals, even in noisy environments and under different operational conditions [8].

Among these methods, Convolutional Neural Networks (CNNs) have proven to be a highly effective tool for automatically extracting key features from raw sensor signals like vibration data, enhancing robustness against various challenges. The vibration behavior of the rocker arm is directly affected by gear mesh excitations, making vibration analysis a key diagnostic approach [9]. CNNs have become the leading choice for mechanical fault diagnosis because they effectively learn local spectral–spatial patterns from vibration or motor current signals. For example, Azamfar et al. [10] used a 2D CNN to combine motor current signals for gearbox fault detection without manual feature engineering. Tang et al. [11] surveyed intelligent diagnosis of helical gearboxes and applied CNN to gear wear images to discriminate wear severity. Other research has focused on lightweight channel attention mechanisms and multi-dimensional attention denoising to boost CNN performance. For example, Cheng et al. [12] proposed a lightweight CNN channel attention mechanism for gearbox diagnosis, reducing model size while maintaining strong accuracy. Liu et al. [13] proposed a multi-dimensional attention denoising CNN that improved robustness to noise and frequency variation in gearbox diagnosis. However, despite their success in capturing local correlations, these CNN models primarily focus on extracting spatial features and often struggle with modeling long-range temporal dependencies in sequential vibration data, which are essential for tracking fault progression and adapting to varying operating conditions. This can lead to overfitting on short-term patterns and inconsistent predictions when analyzing long time series signals [4].

To better manage sequential dynamics, researchers have combined Long Short-Term Memory (LSTM) networks with CNNs. These hybrid models leverage CNNs' feature extraction strengths and LSTMs' ability to model long-range dependencies [14]. LSTM is tailored for sequential data, making it well-suited for tracking how vibration patterns change over time. Research indicates that CNN-LSTM models can outperform others in fault discrimination, particularly in noisy settings [15]. For example, Kang et al. [16] developed a CNN-LSTM model for a wave energy converter gearbox and showed superior discrimination under noisy conditions. Huang et al. [17] proposed a CNN-LSTM-based fault diagnosis method to account for the time delay in fault occurrence and achieved higher diagnostic accuracy and noise robustness. However, these models often follow a simple sequential pipeline, where the temporal modeling capability of the LSTM network is limited by the CNN output, which can reduce the overall diagnostic performance.

Additionally, as networks grow deeper, they become more prone to overfitting, calling for further enhancements [4].

Recent advances in gear fault diagnosis have explored physics-informed hybrid models that integrate domain knowledge with deep learning to improve interpretability and robustness. Zhang et al. [18] proposed CNN-IPINN, which incorporates gear dynamics equations into the loss function to extract time-varying meshing stiffness directly from vibration signals. Schulz et al. [19] developed a multi-task transformer that simultaneously estimates shaft speed and gear mesh frequency alongside fault classification, leveraging physical knowledge of gear meshing phenomena. Li et al. [20] introduced P-ConvSwinFormer, which combines Swin Transformer and CNN with physics-informed feature extraction for noise-robust diagnosis. These physics-informed models demonstrate the value of incorporating domain knowledge into deep learning architectures. However, a critical challenge remains: prediction instability. Transient signal fluctuations and measurement noise can cause these models to produce inconsistent classifications, leading to frequent rapid switching between predicted fault states—a phenomenon termed "classification jitter." This instability undermines the reliability of the diagnostic system, as physical gear faults evolve gradually and do not appear and disappear intermittently. While some studies address noise through separate preprocessing steps, such as Er and Koca [21], who proposed a five-stage hybrid framework integrating EMD-based denoising, detrended fluctuation analysis (DFA), LSTM, and K-means clustering, these methods add complexity and are not integrated into the end-to-end learning process. To address the lack of an integrated mechanism within hybrid deep learning models to enforce temporal consistency, thereby mitigating prediction instability caused by noise without relying on separate denoising stages, this study introduces a Markov Transition Model (MTM) as a probabilistic smoothing layer following a CNN-LSTM feature extractor.

Based on these observations, this study proposes an intelligent vibration measurement-based fault diagnosis framework for shearer rocker arm gears by integrating a CNN, LSTM, and MTM. The proposed framework utilizes a CNN to extract fault features and a LSTM to capture temporal dependencies among vibration segments to model the dynamic evolution of gear fault states. MTM is introduced to refine the accurate classification results by integrating probabilistic transition relationships between shearer rocker arm gear fault states and improve the reliability of gear fault identification under complex operating conditions and noise, which contributes to improved operational safety, reduced maintenance costs, and enhanced reliability of underground mining equipment.

The main contributions of this study include the following:

1. A vibration-based intelligent condition monitoring method is developed for fault diagnosis of shearer rocker arm gears. Unlike traditional signal processing techniques, the proposed model can simultaneously learn local features and global relationships within the dataset, enabling integrated feature extraction, temporal modeling, and probabilistic state refinement.

2. A hybrid CNN-LSTM framework is developed to learn local fault patterns and global structural relationships while simultaneously modeling temporal dependencies in sequential signals.

3. A Markov Transition Model (MTM) is incorporated to refine diagnostic predictions by modeling probabilistic state transitions, thereby improving stability and reducing misclassifications caused by measurement noise.

4. Experimental results demonstrate that the proposed CNN-LSTM-MTM framework improved diagnostic stability and classification accuracy compared with conventional deep learning models.

The paper is organized as follows: Section 2 discusses the theoretical background of the proposed model; Section 3 describes the methodology; Section 4 presents experimental validation, key results, and discussion; and, finally, Section 5 presents the conclusion of this article.

2. Theoretical Background

2.1. Convolutional Neural Network (CNN)

CNN a form of deep feedforward neural network with weight sharing and local connection features [22], which enables powerful feature extraction and pattern recognition capabilities. It has been widely employed in various domains, such as image recognition. A typical CNN is made up of convolutional layer, a pooling layer, and a fully connected layer, as shown in Figure 1.

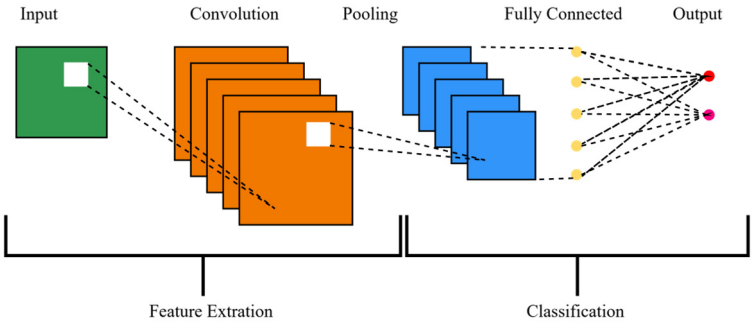


Figure 1. Schematic diagram of the convolution operation.

Convolutional kernels are used in this layer to progressively perform local convolution operations to input features, and the number of convolutional kernels used determines the depth of the output feature map. The convolutional layer function is as follows:

$$x_i^{k+1} = f\left(\sum_j x_j^k \times w_{ij}^{k+1} + b_i^{k+1}\right) \quad (1)$$

where w_{ij}^{k+1} is the convolution kernel with j inputs connecting the ith feature tensor, b_i^{k+1} is the $k+1$ layer bias term, f is the activation function, $\times$ is the convolution operation, x_i^{k+1} is the ith feature tensor output from the layer, and $k+1$, x_j^k is the jth feature tensor output from layer k.

2.2. Long Short-Term Memory (LSTM)

LSTM network, a specialized type of Recurrent Neural Networks (RNNs), excels at handling sequential data. Its internal gating mechanism consists of the forget gate, input gate, and output gate, which control the information flow to prevent the vanishing or exploding gradient problems that often affect traditional RNNs [8]. Its structure is shown in Figure 2. This makes LSTM particularly effective for tasks with time series data, such as vibration signals in fault diagnosis, where capturing temporal dependencies is vital for a precise analysis. By using LSTM, the model is able to process long-sequence signal data and model long-term trend changes in time series data.

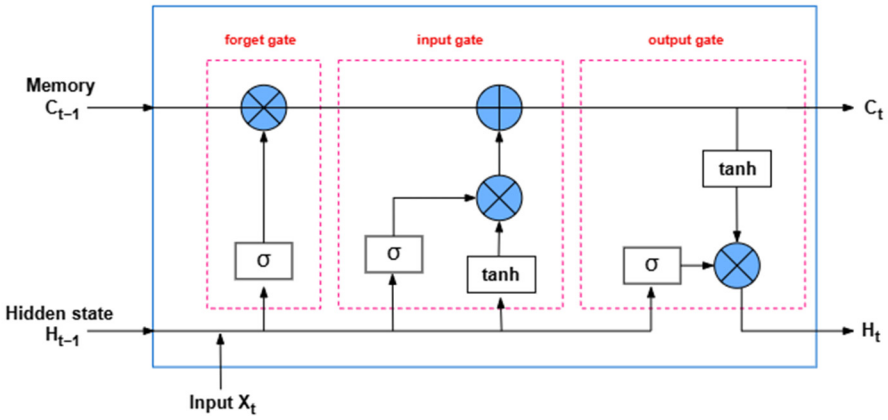


Figure 2. Schematic diagram of LSTM.

The LSTM network includes a cell state (long-term memory) and a hidden state (short-term memory), enabling it to capture dependencies over extended periods. The operations of LSTM units are expressed in Equations (2)–(6).

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \tag{2}$$

$$f_t = \sigma\left(W_f \times [h_{t-1}, x_t] + b_f\right]) \tag{3}$$

$$O_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o) \tag{4}$$

$$\mathbb{C}_t = f_t \times \mathbb{C}_{t-1} + i_t \times \tanh(W_c \times [h_{t-1}, x_t] + b_c) \tag{5}$$

$$h_t = O_t \times \tanh(\mathbb{C}_t) \tag{6}$$

In these equations, i_t, f_t, and o_t are the activation values of the input gate, forget gate, and output gate at time step t, respectively; C_{t-1}, C_t, and h_t denote the previous cell state, the current cell state, and the current hidden state; σ is the activation function; W_i, W_f, W_o, and W_c are the weight matrices corresponding to the input gate, forget gate, output gate, and cell state; $[h_{t-1}, x_t]$ connects the previous hidden state h_{t-1} with the current input x_t into a single vector; b_i, b_f, b_o, and b_C are the bias vectors for the input gate, forget gate, output gate, and cell state. Finally, tanh is the hyperbolic tangent function, and $\tanh(C_t)$ applies it to the current cell state.

2.3. Markov Transition Model (MTM)

MTM offers a probabilistic approach to modeling sequential system behavior by assuming that the next state depends solely on the current one, aligning with the Markov property. This makes the MTM effective for capturing degradation patterns and health transitions in mechanical systems [23]. Unlike a Hidden Markov Model (HMM), which uses hidden states and emission probabilities to model observations, the MTM in this model operates directly on the predicted probability distributions, using only a state transition probability matrix. The specific expression is as follows: the state of the system at time, t, is the probability vector, $\mathrm{p(t)} \in \mathbb{R}^{\mathrm{k}}$, where k is the number of possible states. The evolution of the state distribution is governed by a state transition probability matrix, $T \in \mathbb{R}^{\mathrm{k}\times\mathrm{k}}$, where each element, T_{ij}, is the probability of transitioning from state i to state j in one time step:

$$T_{ij} = P(\text{state} = j \mid \text{previous state} = i) \tag{7}$$

The transition matrix satisfies the stochastic property that each row sums to one:

$$\sum_{j=1}^{k} T_{ij} = 1, \forall i \in \{1, \ldots, k\} \tag{8}$$

Given the current state distribution, $\mathrm{p}(t)$, the distribution at the next time step predicted solely by the Markov chain is:

$$\mathrm{p}(t+1) = \mathrm{p}(t) \cdot T \tag{9}$$

This captures the inherent temporal dependencies in state evolution without requiring detailed modeling of observation emissions. In fault diagnosis applications, this allows the system to enforce realistic progression paths and penalize physically implausible jumps and, when integrated with deep learning, the MTM can refine predictions from neural networks by enforcing temporal consistency.

3. Methodology

A hybrid fault diagnosis framework combining CNN-LSTM-MTM is proposed to ensure stable temporal prediction performance and enhance spatial and temporal feature extraction capability. Its overall framework is illustrated in Figure 3. The model takes the Convolutional Neural Network (CNN) as the core feature extractor, converting raw 1D vibration signals into 2D matrix data to effectively capture local spatial patterns. Given that gear fault diagnosis involves long-term time series data with evolving degradation trends, this study introduces a Long Short-Term Memory (LSTM) network, which has excellent long-term time series feature extraction capability, to model the dynamic evolution of fault states over time. However, LSTM networks alone do not explicitly enforce probabilistic constraints on state transitions, which can lead to physically implausible predictions. To address this limitation, this study further integrates a Markov Transition Model (MTM) to smooth the prediction sequence, penalizing abrupt and unrealistic jumps between fault states, thereby enhancing the diagnostic stability of the model.

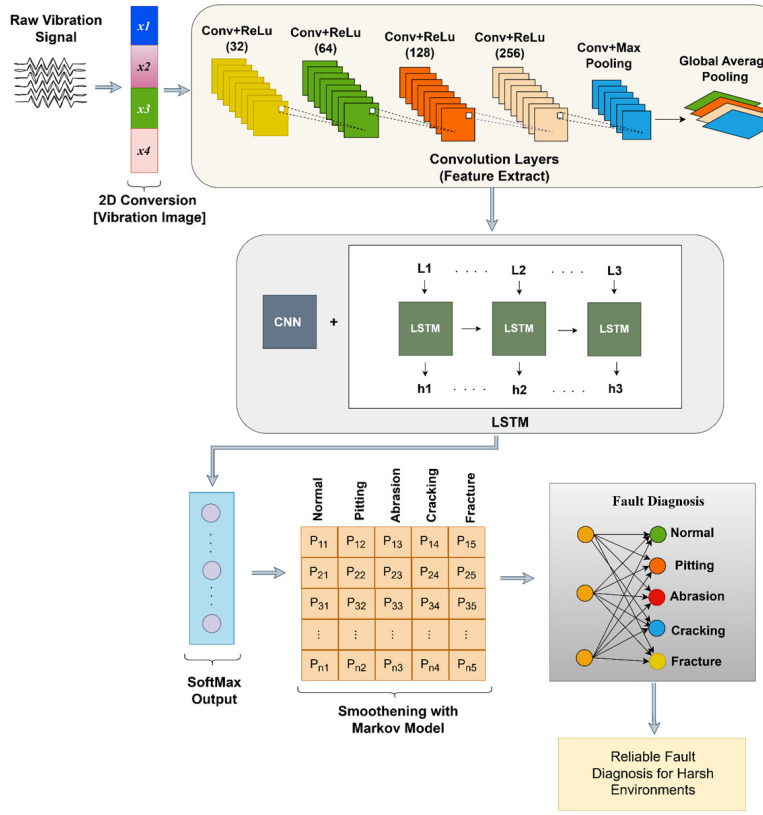


Figure 3. The proposed CNN-LSTM-MTM.

The structure of the CNN-LSTM-MTM integrates multiple components. Its design principle is to improve diagnostic accuracy by combining spatial feature extraction, temporal dependency modeling, and probabilistic state smoothing. The architecture includes the signal input layer, 2D conversion, time-distributed CNN layer, batch normalization and ReLU activation, pooling layers, LSTM layer, and the MTM smoothing layer with a state transition probability matrix. During the study design phase, the model achieved high classification accuracy which is justified by the critical importance of fault diagnosis in underground coal mining operations. For many industrial systems, especially those in safety-critical areas such as mining, energy production, and heavy machinery, accurate fault diagnosis is essential, even if it results in a modest increase in computational complexity.

3.1. Data Preprocessing

In this study, raw vibration signals collected from the gearbox are 1D time series signals. Since the time-distributed CNN and LSTM are effective at processing 2D signals, we converted the 1D vibration signal into a 2D data matrix, as shown in Figure 4.

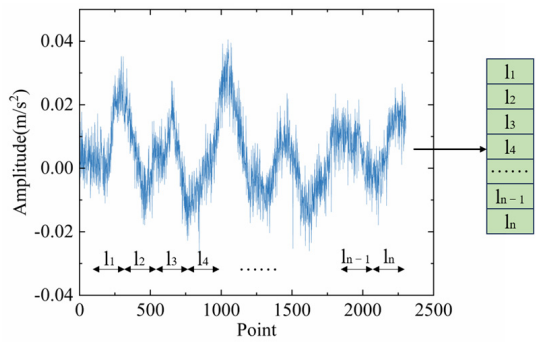


Figure 4. Diagram of data conversion.

The segment length of 4096 points was used to capture data at the 12 kHz sampling rate. A 50% overlap stride was used to maintain temporal continuity between consecutive segments while keeping the total number of samples manageable. The signal segment is set as a sequence of m sampling points recorded as a row vector, l_i, each time and a total of n such row vectors are normalized to form a 2D matrix sample, P, of size $n \times m$. It can be described as follows:

$$P = [l_1, l_2, \ldots, l_n]^T \tag{10}$$

Before constructing P, each signal segment is normalized to the range $[0, 1]$ using min–max normalization to ensure consistent input scaling and improve training stability:

$$x^* = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{11}$$

where x is the original vibration signal, $x_{\min} = 0$ and $x_{\max} = 1$ are the minimum and maximum values within the signal segment, and x^* is the normalized value. This transformation preserves the relative relationships between data points while ensuring all inputs fall within a uniform range.

3.2. CNN for Feature Extraction

The CNN branch is designed to extract local details and global context information from the 2D vibration signal matrix data. The input is expressed as:

$$X \in \mathbb{R}^{64 \times 64 \times 1} \tag{12}$$

which is obtained by reshaping each segmented signal of length 4096 after data splitting into a 2D matrix data. The feature extraction process consists of a sequence of convolutional operations. For the l-th convolutional layer, the output feature map is defined as:

$$\mathrm{F}^{(l)} = \sigma\left(\mathrm{BN}\left(\mathrm{W}^{(l)} \times \mathrm{F}^{(l-1)} + \mathrm{b}^{(l)}\right)\right) \tag{13}$$

where $\times$ is the convolution operation, $W^{(l)}$ and $b^{(l)}$ are the learnable weights and biases of the l-th layer, $\mathrm{BN}(\cdot)$ is batch normalization, $\sigma(\cdot) = \max(0, \cdot)$ is the ReLU activation function, and $F^{(0)} = X$. The network employs four convolutional layers with filter sizes, increasing progressively as 32, 64, 128, and 256, to capture low-level to high-level fault features. Each convolution uses a kernel size of 3×3 with stride of 2×2, which performs implicit downsampling while preserving salient spatial patterns.

After the convolutional blocks, a max pooling operation is applied to further reduce spatial dimensions:

$$\mathrm{F}_{\mathrm{pool}} = \mathrm{MaxPool}\left(\mathrm{F}^{(4)}\right) \tag{14}$$

where a pooling window of 4×4 size is used. To aggregate the spatial feature into a compact representation, a global average pooling (GAP) layer is applied:

$$z_k = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_{pool}^{(k)}(i, j) \tag{15}$$

where z_k is the k-th element of the resulting feature vector, and H and W denote the spatial dimensions of the feature map. The final output of the CNN branch is therefore a feature vector:

$$\mathrm{z} \in \mathbb{R}^{256} \tag{16}$$

which provides a compact representation of the input vibration signal, capturing both local fault details and global context information. This is subsequently passed to the temporal modeling module, LSTM, for sequential dependency learning.

3.3. LSTM for Temporal Sequencing

While the CNN module effectively extracts spatial features from individual vibration segments, the evolution of mechanical faults is fundamentally a temporal process. Therefore, an LSTM branch is fused into the proposed framework to capture the temporal dependencies between vibration segments, as shown in Figure 5.

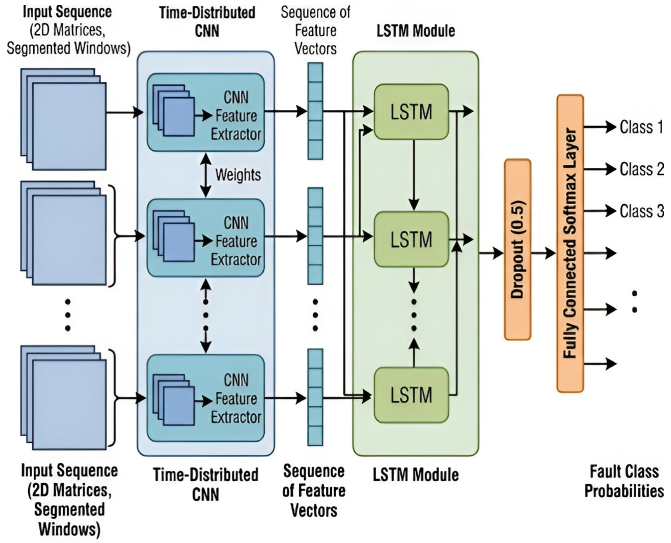


Figure 5. The enhanced structure of the LSTM.

The LSTM branch mainly consists of a time-distributed feature extractor, an LSTM layer, a dropout layer, and a fully connected classification layer. First, the sequence of the vibration matrices is processed using a time-distributed wrapper, which applies the same CNN feature extractor to each time step. This then generates a sequence of compact feature vectors that represent the spatial characteristics of the vibration signal at different time steps. The extracted feature sequence is then fed into the LSTM layer with 128 hidden units, which processes the sequence in order to learn the dynamic evolution of gear fault patterns, which enables the network to capture long-term temporal dependencies in the vibration signals related to the specific fault types and transitions between the different working conditions.

Finally, the regularized feature is passed to a fully connected layer with a softmax activation function, which outputs the probability distribution over all gear fault classes and produces the initial classification result for the input vibration sequence.

3.4. MTM for Smoothing

To further improve the diagnostic stability and predictive accuracy, MTM is added into the framework. Although the CNN-LSTM network provides strong feature extraction and temporal modeling capabilities, prediction results may still fluctuate under noisy conditions. The MTM act as a probabilistic smoothing mechanism to address this issue. It does so by using a state transition probability matrix, T, that captures the gradual and irreversible progression of faults. The transition matrix, T, is constructed based on domain knowledge of gear fault progression for a system with k = 5 fault states: healthy, wear, broken tooth, pitting, and crack. T is a 5×5 matrix where each element $T_{ij} = P(state_j \mid state_i)$ is the probability of transitioning from state i to state j in one time step. The matrix is designed to reflect the following physical principles: (1) transitions are only allowed to the same state or to a more severe state, and regressive transitions are assigned zero probability; (2) diagonal entries are set high (0.85–0.95) to reflect temporal persistence of fault states; (3) direct jumps to non-consecutive states are assigned very low or zero probabilities. Based on these principles, the transition matrix used in this study is defined as:

$$T = \begin{bmatrix} 0.95 & 0.04 & 0.01 & 0.00 & 0.00 \\ 0.00 & 0.90 & 0.08 & 0.02 & 0.00 \\ 0.00 & 0.00 & 0.88 & 0.10 & 0.02 \\ 0.00 & 0.00 & 0.00 & 0.85 & 0.15 \\ 0.00 & 0.00 & 0.00 & 0.00 & 1.00 \end{bmatrix} \tag{17}$$

where rows are the current state, and columns are the next state. The diagonal values reflect the increasing persistence of more severe faults, and off-diagonal values allow gradual progression. These values were empirically tuned using the validation set and remained fixed for all experiments. This matrix is used to refine the prediction sequence, reducing abrupt and physically implausible changes in the diagnosed fault type. The fused prediction probability at time step, t, is defined as:

$$P_{fused}(t) = \alpha P_{CNN\text{–}LSTM}(t) + (1 - \alpha) T P_{CNN\text{–}LSTM}(t - 1) \tag{18}$$

where $P_{\mathrm{CNN-LSTM}}(t)$ is the softmax probability vector output by the CNN-LSTM network at time step, t; $\alpha \in [0, 1]$ is the fusion coefficient that regulates the balance between the direct CNN-LSTM prediction and the Markov-adjusted probability; and $TP_{\mathrm{CNN-LSTM}}(t-1)$ is

the Markov-adjusted contribution based on the previous time step. The final gear state is determined by selecting the class with the highest fused probability:

$$\hat{s}(t) = arg\ \max_i [P_{fused}(t)]_i \tag{19}$$

This ensured that the prediction reflects both the current time step and the long-term transition behavior captured by the MTM. This approach effectively stabilizes the sequence of predictions and enhances reliability in noisy and dynamic operating environments.

3.5. Model Training and Testing Process

To validate the effectiveness of the proposed CNN-LSTM-MTM framework, a training and testing process was established. To prevent data leakage, the original signals were first split into training and testing sets at the signal level after overlapping segmentation was applied to each partition to generate the final samples. This process ensured that the network not only learns feature patterns from the vibration data but also maintains good generalization and stability under varying working conditions. The procedure involved three main steps: data preprocessing, model training, and fault diagnosis, as shown in the flowchart in Figure 6.

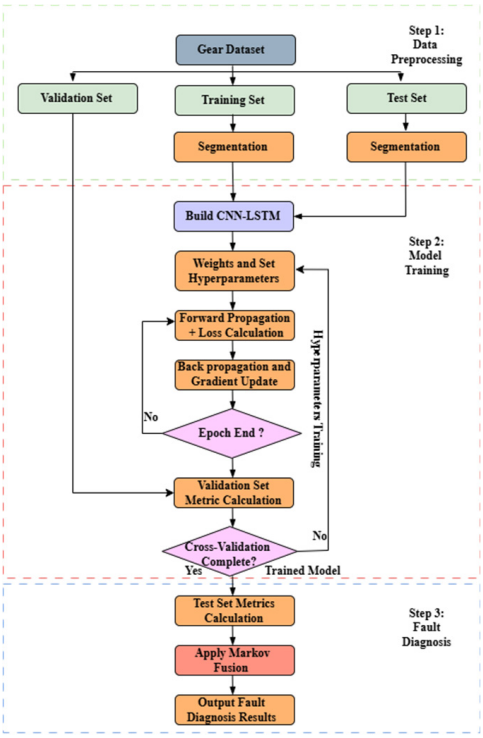


Figure 6. Flowchart of the proposed model.

4. Experimental Validation

To thoroughly assess the effectiveness of the proposed CNN-LSTM-MTM for fault diagnosis in shearer rocker arm gears, a series of comparative experiments was performed. These tests used vibration signals collected from a 2500 kW shearer rocker arm loading test bench at Taizhong Mining Machinery Group's workshop, along with a public dataset provided by Southeast University for verification. In this section, a description of the experimental dataset, the experimental setup and equipment, and the evaluation metrics used for performance assessment, as well as the design of the ablation studies and comparative experiments, were provided.

4.1. Experimental Setting

The proposed network was built using the Keras framework with a TensorFlow back end in Python 3.9. All tests were run on a high-performance workstation featuring an Intel® Core™ i7-10750H CPU @ 2.60 GHz, 16 GB of RAM, and an NVIDIA GeForce RTX 3060 GPU for faster computation. The main hyperparameters of the proposed model configured during training are listed in Table 1.

Table 1. Hyperparameters for the proposed model.

Hyperparameter	Value
Input size	(64, 64, 1)
Optimizer	Adam
Learning rate	0.001
Batch size	64
Epochs	200
LSTM units	128
Dropout rate	0.5
Fusion coefficient (α)	0.7

The input size is $64 \times 64 \times 1$, which corresponds to the reshaped vibration signal segments of length 4096. This process generated 1964 training samples (80%) and 491 testing samples (20%), with 393 samples (20% of the training set) reserved for validation. The model was trained for 200 epochs using the Adam optimizer with categorical cross-entropy loss, with an initial learning rate of 0.001 reduced by a factor of 0.1 every 50 epochs. To prevent overfitting, a dropout rate of 0.5 was applied.

4.2. Evaluation Metrics

Model evaluation is essential in scientific research because it demonstrates the effectiveness of the proposed model. To assess how well the model performs, its evaluation metrics are compared with those of the baseline models. To quantitatively evaluate the proposed model diagnostic performance, four standard classification metrics were employed: accuracy, precision, recall, and F1-score. The mathematical structure of these evaluation metrics is:

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{20}$$

$$Precision = \frac{Number\ of\ true\ positives}{Total\ number\ of\ positive\ predictions} = \frac{TP}{TP + FP} \tag{21}$$

$$Recall = \frac{Number\ of\ true\ positives}{Number\ of\ ture\ positives + Number\ of\ False\ negative} = \frac{TP}{TP + FN} \tag{22}$$

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{23}$$

4.3. Case 1: Taizhong Coal Machinery Dataset

To verify the classification performance of the model, the normal and faulty operating conditions of the spur gear in the high-speed section of the rocker arm transmission system were simulated using the rocker arm loading platform located in the mining workshop of Taizhong Mining Machinery Group, as shown in Figure 7.



Figure 7. Loading test bench of the shearer rocker arm.

Table 2 details the parameters of the load motor, with the rated power of 1000 kW, the rated voltage of 1140 V, the rated current of 506 A, the rated frequency of 50 Hz, and the rated speed of 1485 r/min. The experimental process is as follows. First, the loaded double rocker arm was secured to the bracket, and the output end of the accompanying rocker arm was connected and fixed to the input end of the load motor using the connecting assembly. The load motor was subsequently linked to the load inverter. An acceleration sensor was mounted on the housing near the spur gear in the high-speed section of the rocker arm and connected to the vibration signal acquisition system. The acquisition system was configured with a sampling frequency of 12 kHz.

Table 2. Main parameters of the load motor.

Type	Parameters
Rated power	1000 KW
Rated voltage	1140 V
Rated current	506 A
Rated frequency	50 Hz
Rated speed	1485 r/min

After the installation of the loading platform and the test system, power was supplied to the motors within the test rocker arm. The load motor was controlled through the load inverter to achieve specified loading ratios. During the experiment, loading levels were set at 25%, 50%, and 75%, with durations of 0.5 and 1 h for each level. Under these conditions, vibration signals of the rocker arm spur gears were recorded for five states: normal, pitting, abrasion, cracking, and fracture.

Finally, vibration data corresponding to the 50% loading condition for each gear state in the high-speed zone were selected. The physical gears in both normal and faulty states are shown in Figure 8.

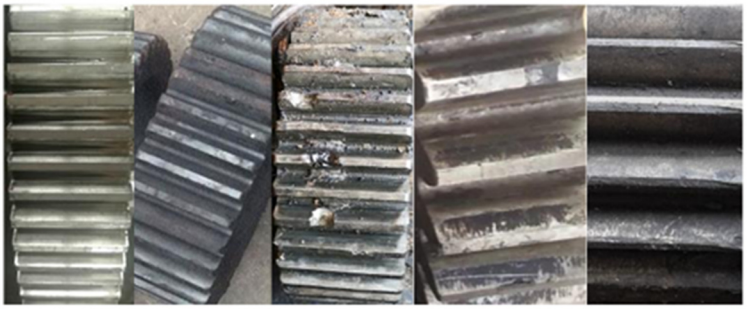


Figure 8. Healthy gear alongside four types of gear faults.

4.3.1. Ablation Experiment

An ablation study was conducted to evaluate the role of each component of the proposed CNN-LSTM-MTM. The full model's performance was compared against several baselines, which include a standalone CNN, a standalone LSTM, a hybrid CNN-LSTM, a CNN-MTM, and an LSTM-MTM. Each experiment was repeated five times with different random initializations, and the performance results are summarized in Table 3 and displayed in Figure 9.

Table 3. Ablation experiment results.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN	93.37	93.41	93.35	93.36
LSTM	81.25	81.53	81.22	81.29
CNN-LSTM	91.11	91.18	91.09	91.10
CNN-MTM	95.82	95.85	95.80	95.81
LSTM-MTM	84.67	84.91	84.65	84.70
Proposed (CNN-LSTM-MTM)	99.79	99.80	99.79	99.79

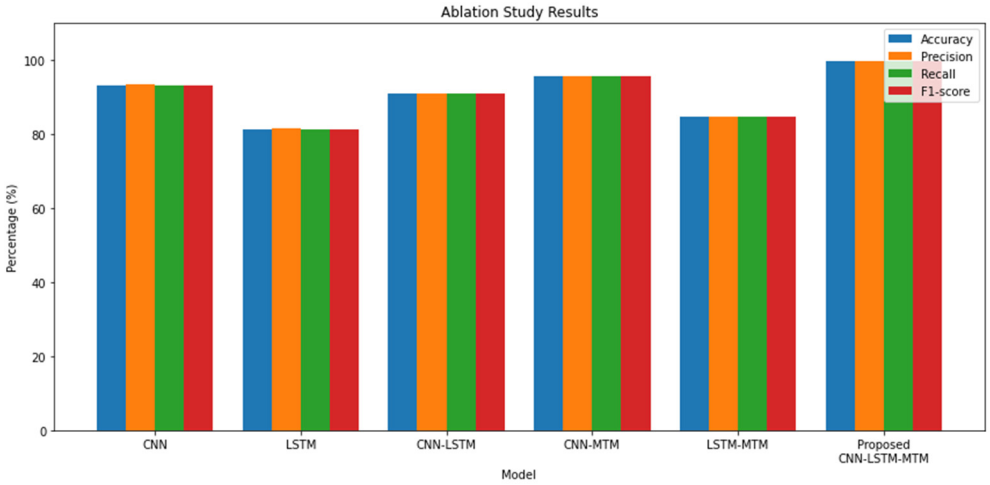


Figure 9. Experimental results of various models within the proposed model.

The results clearly demonstrate the limitations of standalone models and the benefits of the proposed hybrid model. It can be seen that the CNN model achieves a respectable 93.37% accuracy, indicating strong feature extraction, but it is limited by its inability to model temporal evolution. Equally, the standalone LSTM achieves an accuracy of 81.25%, showing that temporal modeling alone is insufficient without spatial feature extraction from the raw signal. The hybrid models show much improvement, with the CNN-LSTM model reaching an accuracy of 91.11%, validating the benefit of combining spatial feature extraction with temporal dependency modeling. Introducing the MTM for smoothing also provided a significant improvement, with the CNN-MTM achieving an accuracy of 95.82%. However, the proposed model significantly outperformed all baseline models, achieving a diagnostic accuracy of 99.79%. This is a 6.42% improvement over the standalone CNN and a 4.14% improvement over the hybrid CNN-MTM, confirming the effect of integrating all three components.

4.3.2. Parameter Optimization Verification

In model construction, the hyperparameter selection plays an important role in the accuracy of the final fault identification process. For example, the fusion parameter, α, in

the MTM module balances the direct CNN-LSTM prediction with the MTM's temporal smoothing. We assessed the proposed model's performance across α values from 0.5 to 1.0. Values below 0.5 were excluded after preliminary experiments showed excessive smoothing, which delayed response to fault and reduced overall accuracy. As shown in Figure 10, as α increases from 0.5, accuracy improves steadily, reaching a peak at 0.7. This indicated that at a lower α value, excessive smoothing suppressed subtle fault signals, lowering accuracy, while at a higher α value, insufficient smoothing leaves the model vulnerable to noise induced fluctuations. The best balance between CNN-LSTM feature extraction and MTM smoothing occurred at $\alpha = 0.7$, leveraging enough historical context to filter noise while remaining responsive to genuine fault states. The MTM reduced noise in raw predictions by enforcing temporal consistency, aligning with the gradual nature of gear fault development. Beyond 0.7, performance slightly stabilized, indicating that too little smoothing (higher α) can interfere with detecting subtle changes, and too much smoothing (lower α) can suppress genuine fault signals. The high accuracy across all α values demonstrates the model's robustness. Therefore, α was set to 0.7 for all subsequent experiments.

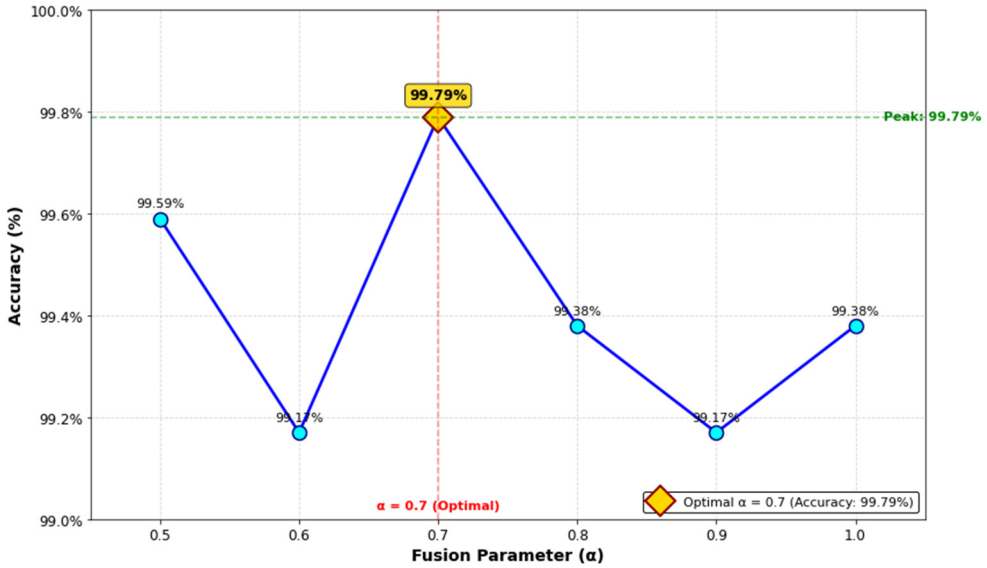


Figure 10. Performance comparison across different values of the parameter (α).

4.3.3. Model Diagnostic Performance Verification

The proposed model was evaluated against several traditional and advanced fault diagnostic models, including a standard CNN [24], ResNet-18 [25], and CNN-LSTM-Attention [8]. The average results from the five repeated experiments are shown in Table 4 and Figure 11, respectively. The proposed model achieved a superior performance across all evaluation metrics, attaining an accuracy of 99.79%, precision of 99.80%, recall of 99.79%, and F1-score of 99.79%. This is a significant improvement of 1.03% over CNN-LSTM-Attention, which attained an accuracy of 98.76%; a 3.25% increase over ResNet-18, which achieved an accuracy of 96.54%; and a 6.42% gain over the standard CNN, which achieved an accuracy of 93.37%. The consistent performance across all metrics demonstrates the effectiveness of the proposed method.

The improvement from CNN to ResNet-18 to CNN-LSTM-Attention to the proposed model suggest the evolution of fault diagnosis capabilities. CNN established strong spatial feature extraction from the 2D vibration signal; ResNet-18 enabled deeper feature extraction through residual connections, capturing more complex patterns; CNN-LSTM-Attention added temporal modeling and feature weighting, allowing the model to focus on the most useful parts of the signal; and the proposed model further enhanced stability

through MTM-based probabilistic smoothing, eliminating false classifications and ensuring temporal consistency.

Table 4. Diagnostic performance of different models.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN [24]	93.37	93.41	93.35	93.36
ResNet-18 [25]	96.54	96.58	96.52	96.53
CNN-LSTM-Attention [8]	98.76	98.79	98.75	98.76
Proposed (CNN-LSTM-MTM)	99.79	99.80	99.79	99.79

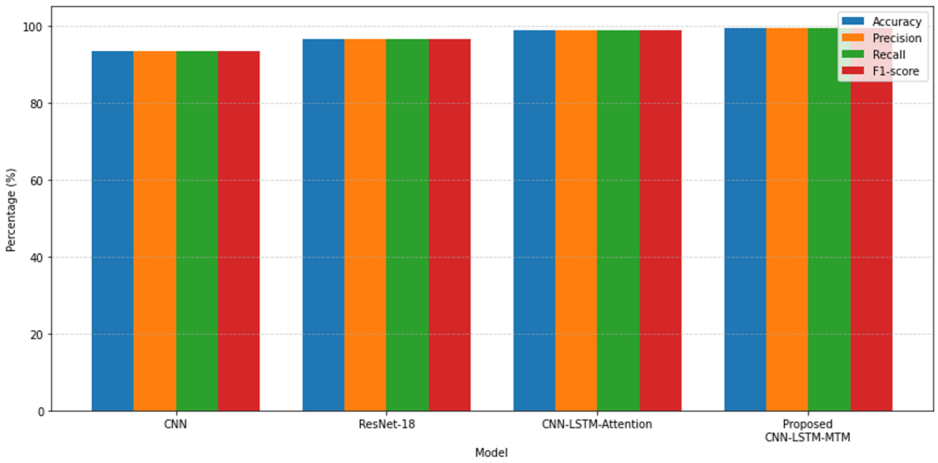


Figure 11. Bar chart comparing different models.

The confusion matrix in Figure 12 shows a detailed view of classification performance of the proposed model, showing good classification results. The only misclassification occurred between physically similar fault types: wear and healthy. No misclassifications occurred between physically dissimilar states, for example healthy and crack or broken tooth and pitting, indicating that the model respects the fault progression structure. All broken tooth, pitting, and crack samples were perfectly classified, further confirming the model's robust capability for identifying distinct fault types.

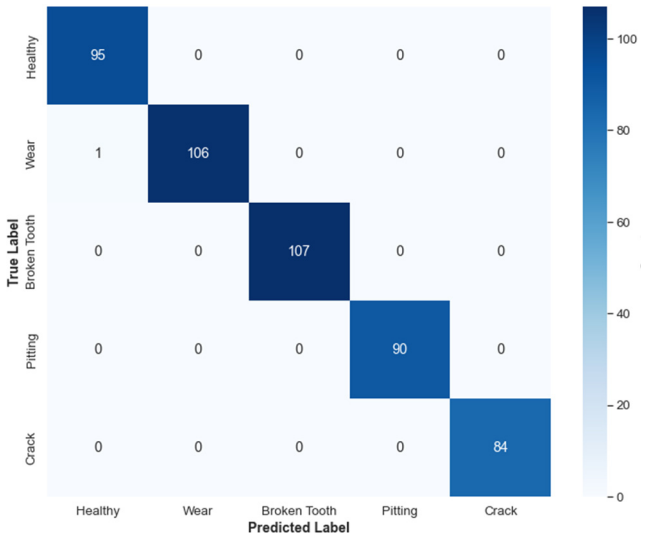
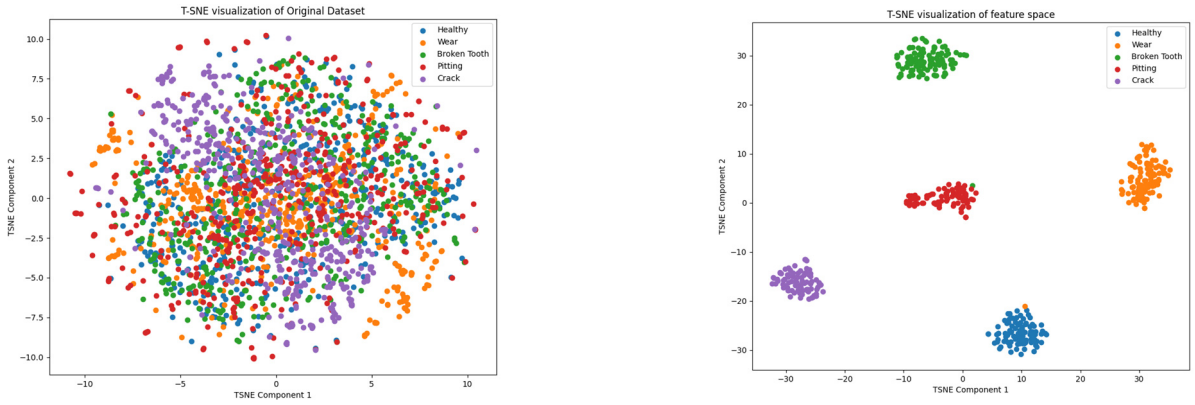


Figure 12. Confusion matrix.

To further analyze the effectiveness of the proposed model, a T-SNE visualization analysis plot was used to see the superiority of the model in terms of its classification across gear fault types. Figure 13 shows the original clustering of samples before classification and the separations between different fault conditions, showing how the data points related to each fault condition are distributed in the feature space.



(**a**) (**b**)

Figure 13. (**a**) Original T-SNE plot. (**b**) T-SNE plot of the results after classification.

4.3.4. Noise Resistance Test

The vibration signals collected from gearboxes in real industrial environments are frequently contaminated by noise, which can easily obscure key fault features affecting identification accuracy. Therefore, to ensure the performance of the proposed fault diagnosis method under strong noise conditions is necessary. In the test, to simulate the noisy environment, Gaussian white noise is added to the original signal, generating datasets with signal-to-noise ratio (SNR). The SNR is defined as:

$$SNR_{dB} = 10log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) \tag{24}$$

where P_{signal} and P_{noise} are the powers of the original signal and the added noise, respectively. In this article, the SNR levels of 5 dB, 0 dB, and −5 dB were evaluated. The accuracy of the proposed model and several other models, such as CNN [24], ResNet-18 [25], CNN-LSTM-Attention [8], under these noisy conditions are compared in Figure 14 and detailed in Table 5.

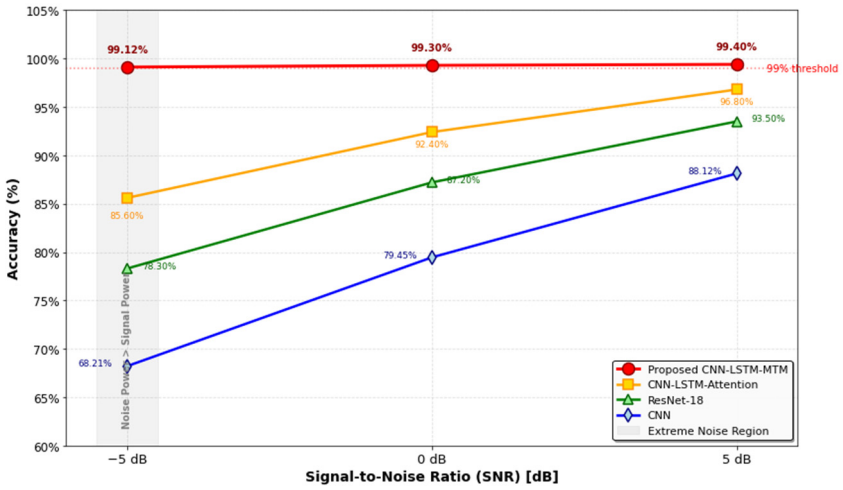


Figure 14. Results comparison of the accuracy of different models under noise.

Table 5. Accuracy (%) of the model under varying SNR levels.

Model	5 dB	0 dB	−5 dB
CNN [24]	88.12	79.45	68.21
ResNet-18 [25]	93.50	87.20	78.30
CNN-LSTM-Attention [8]	96.80	92.40	85.60
Proposed (CNN-LSTM-MTM)	99.40	99.30	99.12

The results from Figure 14 show that the proposed model demonstrates excellent noise robustness, maintaining an accuracy above 99% across all SNR levels, with only a 0.33% degradation from 5 dB 99.40% to −5 dB 99.12%. In contrast, the other models show significant degradation as noise increased. The CNN accuracy decreased from 88.12% at 5 dB to 68.21% at −5 dB, a 19.91% reduction. ResNet-18 declined from 93.50% to 78.30%, a 15.20% reduction, and CNN-LSTM-Attention model dropped from 96.80% to 85.60%, a 11.20% reduction. Notably, while the CNN-LSTM-Attention model provided some noise resilience with an 11.20% drop compared to CNN's 19.91% reduction, it still falls far short of the performance achieved by the proposed model. At the most challenging noise level of −5 dB, where noise exceeds signal, the proposed model achieved 99.12% accuracy, an improvement of 30.91% over CNN, 20.82% over ResNet-18, and 13.52% over CNN-LSTM-Attention.

These results show that the proposed MTM plays a critical role in enhancing diagnostic reliability under the severe noise conditions by smoothing the CNN-LSTM output sequence with a state transition matrix that penalizes noise-induced fluctuations. The MTM effectively filters out transient errors while preserving genuine fault signals. This capability, combined with the CNN feature extraction and the LSTM temporal modeling enabled the CNN-LSTM-MTM method to maintain high accuracy even at −5 dB SNR, showing strong anti-noise performance and the ability to adapt to harsh working conditions.

4.4. Case 2: Southeast University Gear Dataset

To assess the proposed model's ability to generalize and improve the convincing power, we evaluated it using the public gearbox dataset provided by Southeast University (SEU). The experimental setup and fault types are shown in Figure 15, which comprise a motor controller, motor, planetary gearbox, parallel gearbox, and brake. The model 608A11 vibration sensor was selected to collect the x-axis vibration signals of the gearbox under two working conditions (20 Hz_0 V and 30 Hz_2 V) and a sampling frequency of 5120 Hz. The gear under each working condition has five states recorded namely chipped, miss, root, surface, and healthy in Table 6, and the division of the dataset for each working condition shown in Table 7. To ensure impartiality in the comparison and avoid result variability, we repeated the test process five times and used the average values.

Table 6. Gearbox fault type description.

Label	Description	Type
0	No damage	Health
1	Crack occurs in the gear foot	Chipped
2	Missing one foot in the gear	Miss
3	Crack occurs in the root of the foot	Root
4	Wear occurs in the surface of the gear	Surface

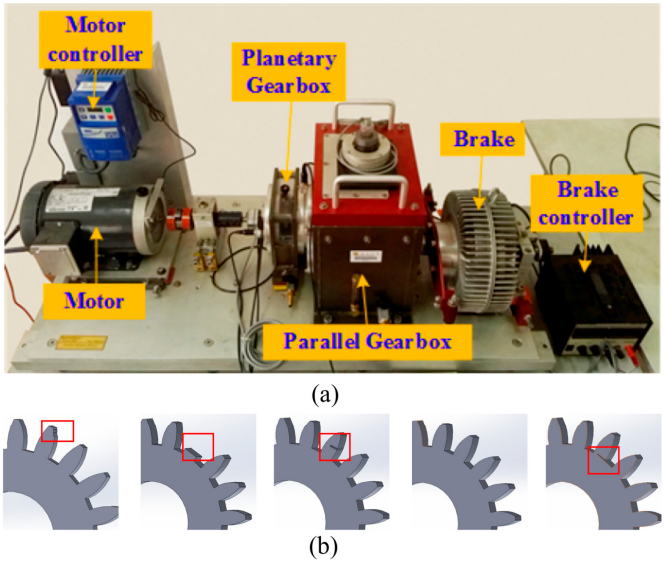


Figure 15. (**a**) Diagram of the gearbox test rig and (**b**) show five different gear fault types.

Table 7. Division of dataset.

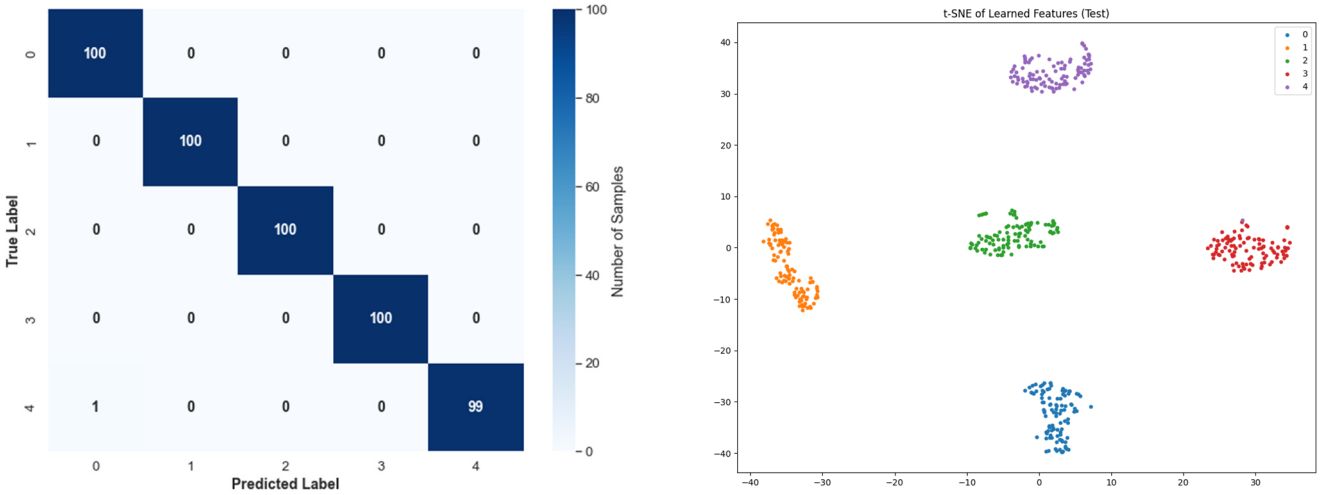
Label	Fault Type	Training	Validation	Test
0	Health	60	20	20
1	Chipped	60	20	20
2	Miss	60	20	20
3	Root	60	20	20
4	Surface	60	20	20

Figure 16 presents the confusion matrix and T-SNE plot for the 20 Hz_0 V working condition, respectively. As shown in Figure 16a, the proposed model correctly classified 499 out of 500 samples correctly, achieving an accuracy of 99.8%, with only one surface fault sample misclassified as root and all other classes are identified without error. Figure 16b displays the T-SNE plot of the five fault states clearly separated. Similarly, Figure 17 presents the confusion matrix and T-SNE plot for the 30 Hz_2 V working condition. As shown in Figure 17a, the model correctly classifies 498 out of 500 samples correctly, achieving an accuracy of 99.6%, with two misclassifications including one missing tooth sample misclassified as surface, one root fault sample misclassified as surface, and all other classes identified without error. Figure 17b displays the T-SNE plot of the five fault states also clearly separated. Therefore, the proposed method shows strong feature extraction capability and effective fault diagnostic performance.

To further validate the diagnostic capability of the proposed model, we compared its performance with several state-of-the-art methods. The results are summarized in Table 8.

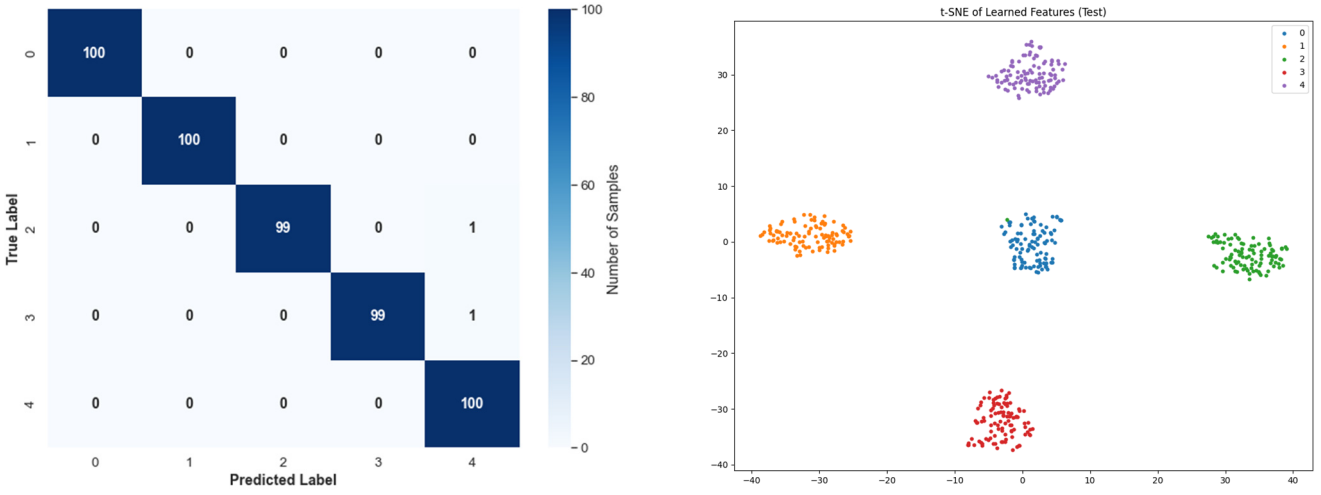
Table 8. Comparison with different methods.

Method	Accuracy (20 Hz_0 V)	Accuracy (30 Hz_2 V)	Average Accuracy
Deep Random Forest [26]	95.4%	94.8%	95.1%
DCNN [2]	98.1%	97.5%	97.8%
WDCNN [27]	99.2%	98.6%	98.9%
Swin-CNN-GAM [28]	99.55%	99.41%	99.48%
Proposed (CNN-LSTM-MTM)	99.8%	99.6%	99.7%



(**a**) (**b**)

Figure 16. (**a**) Confusion matrix, (**b**) T-SNE of 20 Hz_0 V.



(**a**) (**b**)

Figure 17. (**a**) Confusion matrix, (**b**) T-SNE of 30 Hz_2 V.

As shown in Table 8, Deep Random Forest model uses vibration signals for fault diagnosis, achieving an average accuracy of 95.1%. Despite using only single-channel vibration data, the proposed model outperformed this multi-sensor approach by 4.6%, demonstrating the effectiveness of the hybrid CNN-LSTM-MTM architecture. DCNN achieved 97.8% average accuracy using vibration-only input, while WDCNN further improved to 98.9% through the use of wide first-layer kernels designed for raw vibration signal processing. Swin-CNN-GAM, a recent dual-branch network combining Swin Transformer and CNN with a global attention module, achieved 99.48% average accuracy on the SEU dataset. The proposed model achieved the highest average accuracy of 99.7%, with 99.8% accuracy at 20 Hz_0 V and 99.6% at 30 Hz_2 V. This is a 0.22% improvement over Swin-CNN-GAM and a 0.8% improvement over WDCNN. These results show the feature extraction and temporal modeling capabilities of the proposed model, further enhanced by MTM probabilistic smoothing. Consistent performance across both working conditions confirms the model's diagnostic ability and establishes the CNN-LSTM-MTM framework as a competitive approach for gear fault diagnosis.

5. Conclusions

To address the limitations of existing diagnostic methods in handling nonlinear and noisy vibration signals in underground mining environments, this study proposed a hybrid CNN-LSTM-MTM method for vibration-based fault diagnosis of shearer rocker arm gears. The proposed method integrates a Convolutional Neural Network for spatial feature extraction, a Long Short-Term Memory network for temporal dependency modeling, and a Markov Transition Model for probabilistic smoothing of diagnostic results. Experimental verification was conducted using gear vibration signals collected from the rocker arm platform of the coal mining machine at Taizhong Coal Machinery and the public Southeast University gearbox dataset.

1. The hybrid CNN-LSTM framework effectively combined spatial feature extraction and temporal sequence modeling, enabling the model to capture both local vibration characteristics and the dynamic evolution of gear fault states. This improved the diagnostic ability of the method, achieving a classification accuracy of 99.79% on the Taizhong Coal Machinery dataset.

2. The introduction of the Markov Transition Model improved prediction stability by incorporating probabilistic state transition relationships between gear health conditions compared with models without the MTM component. The proposed framework produced smoother and more consistent classification results by reducing abrupt and physically implausible state transitions.

3. The proposed framework demonstrated strong diagnostic results; when evaluated on the public Southeast University gearbox dataset, the model achieved 99.7% accuracy, confirming its ability to identify multiple gear fault types.

4. The model also exhibited strong robustness to noise, with experimental results showing that the proposed model-maintained accuracy above 99% even under severe noise condition of -5 dB SNR, indicating its suitability for condition monitoring in industrial environments such as underground coal mines.

Although the proposed method achieved promising results, several limitations remain that require further investigation.

1. Limitation of Single Sensor Data: The current diagnostic framework relies primarily on vibration signals. In future work, multisource sensor data such as acoustic emission, temperature, and current signals will be incorporated to improve diagnostic reliability and enhance the robustness of the monitoring system.

2. Difference Between Experimental and Real Operating Conditions: The experimental validation was performed using test platform data under controlled conditions. However, real mining environments involve more complex operating conditions, including variable loads, speed fluctuations, and non-stationary, impulsive noise that differs from the simulated Gaussian white noise used in this study. Future research will focus on validating the proposed framework using real operational data from working shearers and improving the adaptability of the model under varying operating conditions.

3. Generalization Claims: While the proposed model achieved strong performance on the public Southeast University gearbox dataset, this validation was conducted under two specific operating conditions. Generalization to other machinery types, fault conditions, or operating environments requires further validation. Future work will focus on evaluating the model on additional datasets and under more diverse operating conditions to establish broader applicability.

https://doi.org/10.3390/electronics15081587

Author Contributions: J.F. worked the manuscript writing, model design, and coding; X.S. worked on coding and provided the datasets; G.S.G. was responsible for conceptualization and editing; H.D. supervised and evaluated the methods used for the study; N.L. was responsible for correcting every technical and grammatical error in the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: The authors gratefully acknowledge the support from the following foundations: National Natural Science Foundation of China (No. 52574201), Shanxi Province Foreign Experts Project (No. 202504051012018), and Key Research and Development Project (No. 202402100101005).

Data Availability Statement: The data used and analyzed during the current study are not publicly available for legal/ethical reasons but are available from the corresponding author on reasonable request.

Acknowledgments: We sincerely appreciate the faculty members in the College of Mechanical Engineering and the College of Computer Science at Taiyuan University of Technology, China, for their valuable support and guidance.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Yang, X.; Chen, H.; Mao, J.; Wei, Y. Dynamical Behavior of Coal Shearer Traction-swing Coupling under Corrected Loads. *Sci. Rep.* **2020**, *10*, 8630. [CrossRef]
2. Janssens, O.; Slavkovikj, V.; Vervisch, B.; Stockman, K.; Loccufier, M.; Verstockt, S.; Van de Walle, R.; Van Hoecke, S. Convolutional Neural Network Based Fault Detection for Rotating Machinery. *J. Sound Vib.* **2016**, *377*, 331–345. [CrossRef]
3. Ding, H.; Yang, L.; Yang, Z. A Predictive Maintenance Method for Shearer Key Parts Based on Qualitative and Quantitative Analysis of Monitoring Data. *IEEE Access* **2019**, *7*, 108684–108702. [CrossRef]
4. Jia, H.; Qin, D. Effects of ranging arm housing's large deformation on the dynamic responses of the cutting transmission system for a drum shearer. *Adv. Mech. Eng.* **2019**, *11*, 1687814019844361. [CrossRef]
5. Zhang, Y.-M.; Wang, T.; Huang, J. Fatigue reliability-based sensitivity design of planet gear for shearer rocker arm system. *J. Northeast. Univ. (Nat. Sci.)* **2016**, *37*, 1426–1431. [CrossRef]
6. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
7. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
8. Cheng, K.; Cheng, L.; Mao, T.; Wu, M. Fault Diagnosis of Gearbox Based on CNN and LSTM with Attention Mechanism. *Highlights Sci. Eng. Technol.* **2024**, *93*, 285–294. [CrossRef]
9. Zhu, L.; Li, H.; Lin, Z.; Chen, H.; Wang, W.; Han, H. *Analysis of the Inherent Characteristics of the Power Transmission System of the Shearer's Rocker Arm*; IEEE: New York, NY, USA, 2020; pp. 47–50. [CrossRef]
10. Azamfar, M.; Singh, J.; Bravo-Imaz, I.; Lee, J. Multisensor data fusion for gearbox fault diagnosis using 2-D convolutional neural network and motor current signature analysis. *Mech. Syst. Signal Process.* **2020**, *144*, 106861. [CrossRef]
11. Tang, X.; Xu, Y.; Sun, X.; Liu, Y.; Jia, Y.; Gu, F.; Ball, A.D. Intelligent fault diagnosis of helical gearboxes with compressive sensing based non-contact measurements. *ISA Trans.* **2023**, *133*, 559–574. [CrossRef]
12. Cheng, X.; Dou, S.; Du, Y.; Wang, Z. Gearbox fault diagnosis method based on lightweight channel attention mechanism and transfer learning. *Nat. Portf.* **2024**, *14*, 743. [CrossRef]
13. Yu, T.; Jiang, Z.; Ren, Z.; Zhang, Z.; Zhang, Y.; Zhou, S.; Zhou, X. A convolutional multisensor fusion fault diagnosis framework based on multidimensional distance matrix for rotating machinery. *Struct. Health Monit.* **2024**, *25*, 468–483. [CrossRef]
14. Zhang, J.; Wan, C.; Wang, J.; Chen, C.; Wang, T.; Zhang, R.; Guo, H. Research on the 'shape-performance-control' integrated digital twin system for boom-type roadheaders. *Sci. Rep.* **2024**, *14*, 5780. [CrossRef] [PubMed]
15. Jiang, Y.; Zhang, X.; Zhang, C.; Li, Z.; Sheng, C. Progress on Fault Mechanisms for Gear Transmissions in Coal Cutting Machines: From Macro to Nano Models. *J. Nanosci. Nanotechnol.* **2017**, *17*, 2274–2284. [CrossRef]
16. Kang, J.; Zhu, X.; Shen, L.; Li, M. Fault diagnosis of a wave energy converter gearbox based on an Adam optimized CNN-LSTM algorithm. *Renew. Energy* **2024**, *231*, 121022. [CrossRef]
17. Huang, T.; Zhang, Q.; Tang, X.; Zhao, S.; Lu, X. A novel fault diagnosis method based on CNN and LSTM and its application in fault diagnosis for complex systems. *Artif. Intell. Rev.* **2021**, *55*, 1289–1315. [CrossRef]
18. Zhang, Z.; Wan, C.; Li, L.; Pang, Y.; Xiong, G.; Zhang, C.; Fan, X.; Wang, Y. CNN-IPINN: A method study on applying physics-informed neural networks to gear fault diagnosis. *Meas. Sci. Technol.* **2025**, *36*, 106101. [CrossRef]
19. Yoshimatsu, O.; Knoll, E.; Sendlbeck, S.; Otto, M.; Stahl, K. Transformer for gear fault diagnosis enhancing robustness through physics-informed multi-task learning. *Forsch. Im Ingenieurwesen* **2025**, *89*, 117. [CrossRef]

20. Li, D.; Yang, J.; Xu, W.; Zhu, J. P-ConvSwinFormer: A physics-informed bearing fault diagnosis framework using global-local and physical information under noise interference. *Meas. Sci. Technol.* **2025**, *36*, 116211. [CrossRef]
21. Er, M.B.; Koca, T. A Novel Approach for Motor Bearing Fault Detection Using EMD-Based Denoising and Detrended Fluctuation Analysis & LSTM Multimodal Hybrid Features with K-Means Clustering. *J. Vib. Eng. Technol.* **2026**, *14*, 103. [CrossRef]
22. Zhang, F.; Zhu, Y.; Zhang, C.; Yu, P.; Li, Q. Abnormality Detection Method for Wind Turbine Bearings Based on CNN-LSTM. *Energies* **2023**, *16*, 3291. [CrossRef]
23. Shen, H.; Nie, Z.; Xia, J.; Wang, Q. A Markov method for a mechanical system reliability assessment using discrete degradation data. *J. Phys. Conf. Ser.* **2024**, *2815*, 012048. [CrossRef]
24. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
25. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2016; pp. 770–778. [CrossRef]
26. Li, C.; Sanchez, R.-V.; Zurita, G.; Cerrada, M.; Cabrera, D.; Vásquez, R.E. Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals. *Mech. Syst. Signal Process.* **2016**, *76–77*, 283–293. [CrossRef]
27. Zhang, W.; Peng, G.; Li, C.; Chen, Y.; Zhang, Z. A New Deep Learning Model for Fault Diagnosis with Good Anti-Noise and Domain Adaptation Ability on Raw Vibration Signals. *Sensors* **2017**, *17*, 425. [CrossRef]
28. Huang, X. Bearing Fault Diagnosis Based on LBWO-VMD and Swin-CNN-GAM Dual-Branch Network. *Eng. Res. Express* **2026**, *8*, 065233. [CrossRef]

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.