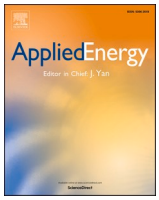
Applied Energy 372 (2024) 123773

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy



A hybrid deep learning model towards fault diagnosis of drilling pump

Junyu Guo [a,b,d,*], Yulai Yang [a,b], He Li [c,d,**], Jiang Wang [a,b], Aimin Tang [e], Daiwei Shan [e], Bangkui Huang [e]

[a] Key Laboratory of Oil & Gas Equipment, Ministry of Education, Southwest Petroleum University, Chengdu, Sichuan, 610500, PR China
[b] School of Mechatronic Engineering, Southwest Petroleum University, Chengdu, Sichuan, 610500, PR China
[c] School of Engineering, Liverpool John Moores University, Liverpool, 3 Byrom Street, L3 3AF, UK
[d] Centre for Marine Technology and Ocean Engineering (CENTEC), Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal
[e] Sichuan Honghua Petroleum Equipment Co., Ltd, Guanghan, Sichuan, 618300, PR China

HIGHLIGHTS

- A novel intelligent drilling pumps fault diagnosis method, WCCN-BiLSTM, is proposed.
- The WCCN is constructed for noise reduction and feature extraction of signals
- The CBAM module embedded in WCCN enhances the feature representation of channel and spatial dimension.
- The diagnostic ability and generalization ability of the model are verified by drilling pump data under different working conditions.

ARTICLE INFO

Keywords:
Drilling pump
Fault diagnosis
WaveletKernelNet-CBAM net
Bidirectional long-short term memory

ABSTRACT

This paper proposes a novel method namely WaveletKernelNet-Convolutional Block Attention Module-BiLSTM for intelligent fault diagnosis of drilling pumps. Initially, the random forest method is applied to determine the target signals that can reflect the fault characteristics of drilling pumps. Accordingly, the WaveletKernelNet-Convolutional Block Attention Module Net is constructed for noise reduction and fault feature extraction based on signals. The Convolutional Block Attention Module embedded in WaveletKernelNet-CBAM adjusts the weight and enhances the feature representation of channel and spatial dimension. Finally, the Bidirectional Long-Short Term Memory concept is introduced to enhance the ability of the model to process time series data. Upon constructing the network, a Bayesian optimization algorithm is utilized to ascertain and fine-tune the ideal hyperparameters, thereby ensuring the network reaches its optimal performance level. With the hybrid deep learning model presented, an accurate fault diagnosis of a real five-cylinder drilling pump is carried out and the results confirmed its applicability and reliability. Two sets of comparative experiments validated the superiority of the proposed method. Additionally, the generalizability of the model is verified through domain adaptation experiments. The proposed method contributes to the safe production of the oil and gas sector by providing accurate and robust fault diagnosis of industrial equipment.

1. Introduction

Oil and gas are still powering the world as it has been in last centuries, in spite of the surge demand and development of renewable energies, such as wind, hydropower, solar, nuclear, and so forth [1–3]. Different from energy production equipment operate on land, drilling devices works underground or under the sea, representing low accessibility, harsh working condition, and limited intervention with operators, maintenance crew, and other stakeholders [4–6].

The drilling pump, one of the riskiest components among others of drilling devices, transports drilling mud to the bottom of wells through the high-pressure pipeline and the center hole of the drill pipe to maintain borehole cleanliness, support lubrication and cooling to the drill bit, and discharge the rock chips produced during drilling to the surface [7]. Failed to promptly identify a fault in a drilling pump can result in equipment damage, causing delays in field operations and resulting in property damage. Fault diagnosis of drilling pumps is

* Corresponding author at: School of Mechatronic Engineering, Southwest Petroleum University, Chengdu, Sichuan, 610500, PR China.
** Corresponding author at: School of Engineering, Liverpool John Moores University, Liverpool, 3 Byrom Street, L3 3AF, UK.
E-mail addresses: junyguo@163.com (J. Guo), he.li@centec.tecnico.ulisboa.pt (H. Li).

https://doi.org/10.1016/j.apenergy.2024.123773
Received 23 October 2023; Received in revised form 12 April 2024; Accepted 20 June 2024
Available online 26 June 2024

Nomenclature

Abbreviations

CNN	Convolutional Neural Network
CWConv	Continuous Wavelet Convolutional
CBAM	Convolutional Block Attention Module
WCCN	WaveletKernelNet-CBAM Net
BiLSTM	Bidirectional Long-short Term Memory
RF	Random Forest
RNN	Recurrent Neural Network
LSTM	Long-short Term Memory
GRU	Gate Recurrent Unit
BO	Bayesian Optimization
BIConvLSTM	Bidirectional-convolutional LSTM
MDRL	Multi-scale Deep Residual Nearning
MSCNN	Multi-scale Convolutional Neural Network
SCN	Sparse Convolutional Network
CWT	Continuous Wavelet Transform
SPM	Strokes per Minute

importance for oil, gas, and other energy drilling operations. The mentioned device accounts a large proportion of the drilling device malfunctions during the production process in land or at the sea according to the field data. The root reasons can be traced back to [8,9]: (i) harmful working conditions including the abrasive and corrosive mud, the high-intensity working demand, and uncertain drilling operation underground or at the sea; (ii) low accessibility of maintenance which formulates that a minor malfunction could results in unrated production and can only be fixed after drilling actions completion. Accordingly, it has been a long-term, urgent, surge demand of the failure identification and prevention of such a device which relies heavily on the state-awareness of drilling pumps. It is, however, still based on the subjective judgment of experienced operators.

To be specific, experienced operators identify the healthy state of drilling pumps according to the sound generated during the drilling process. The usable information given by the manual identification is somehow limited by only providing a pre-judgment on presence and seriousness of malfunctions [10], more detailed and practical decision-making assistance information like the location and remaining useful operation time of drilling pumps are still unavailable according to the manual identification way. Hence, practitioners are seeking automatic state-awareness methods or useful tools to support the optimal and reliable operation of devices with the help of the advance hardware like sensors and applicable methods like deep learning models.

This study presents a novel fault diagnosis framework, called WCCN-BiLSTM, for drilling pumps. The core of this framework is the WaveletKernelNet-CBAM Net (WCCN) model, which is seamlessly integrated with a Bidirectional Long-Short Term Memory (BiLSTM) architecture. An accurate fault diagnosis of a five-cylinder drilling pump is carried out with the assistance of the method. Overall, this method contributes to safe production of the oil and gas sector and benefit for accurate and robust fault diagnosis of industrial equipment.

The remaining sections of the paper are structured as follows: Section 2 offers an overview of the current state-of-the-art and presents the academic perspective on the problem statement. Section 3 details the methodology employed. Section 4 showcases the drilling pump and its fault diagnosis process. Finally, Section 5 presents the conclusion.

2. State of the art and problem statement

2.1. Data-driven fault diagnosis

2.1.1. From CNN to BiLSTM

Traditional data-driven fault diagnosis methods such as machine learning have been implemented to solve real-world engineering problems [11]. For example, Chen et al. [12] utilized a hierarchical machine learning approach to detect gearbox faults employing a two-tier model. Li et al. [13] introduced an extreme learning machine technique for diagnosing faults in hydraulic pumps, specifically addressing the challenge of complex fault diagnosis in axial piston pumps. Meanwhile, Liu et al. [14] presented an enhanced machine learning-based fault diagnosis method for multi-phase drive systems. Lan et al. [15] applied signal processing technology and extreme learning machine to construct an intelligent network for fault diagnosis of slipper wear of axial piston pumps. However, these methods suffer from shortcomings like rely heavily on prior experience and knowledge and have difficulties in extracting deep and nonobvious failure features [16,17].

In recent years, the field of fault diagnosis has witnessed a significant shift towards deep learning techniques, addressing the limitations inherent in traditional methods as previously mentioned [18]. This trend was attributed to the adoption of Convolutional Neural Network (CNN), which excels in identifying spatial features and structures through the use of multiple layers with shared weights [19]. Kumar et al. [20] introduced a centrifugal pump fault detection method using a CNN enhanced by integrating an entropy-based divergence function in its cost function, applying acoustic images processed through analytical wavelet transform, to effectively identify centrifugal pump defects. Similarly, Tang et al. [21] introduced an enhanced CNN model with an adaptable learning rate for hydraulic piston pump fault diagnosis, effectively utilizing vibration, pressure, and sound signals to improve detection accuracy. To realize fault diagnosis of drilling pump, Li et al. [22] developed a CNN-based fault diagnosis method, which utilizes an innovative expansion operation to convert one-dimensional vibration signals into enhanced three-dimensional images. Tang et al. [23] utilized the generalized S transform to convert raw vibration signals into a time-frequency image dataset, followed by the application of a CNN model for fault diagnosis of the drilling pump. Similar to CNN, Recurrent Neural Network (RNN) is a powerful tool for pattern recognition, feature extraction, and sequence modeling. Characterized by their cyclic connections, RNN can process variable-length sequences and pass implicit states at each timestep, a functionality that enables them to effectively capture timing information within these sequences [24]. Wang et al. [25] developed an RNN-based multi-fault diagnosis method for wind power systems, utilizing 30 characteristic parameters like wind speed and rotor speed as inputs. Nevertheless, RNNs face challenges with gradient vanishing or exploding when handling long sequence inputs. To counter this, Hochreiter and Schmidhuber [26] developed the Long Short-Term Memory (LSTM) network, which integrates memory units into the hidden layer, enhancing RNN. LSTM utilizes gated units to control information flow, allowing the model to learn long-term dependencies effectively. Bie et al. [27] developed a fault diagnosis method for reciprocating pumps using LSTM, enhanced by the Improved CEEMDAN algorithm. The approach utilizes vibration signals, effectively extracting fault features for classification. Additionally, Chung et al. [28] introduced the Gated Recurrent Unit (GRU) network, an optimization of the LSTM structure. This approach simplifies the connection architecture and reduces the number of trainable parameters, thus boosting training efficiency while maintaining the memory functionality within the network. Miao et al. [29] developed a planetary gearbox fault diagnosis method based on GRU, enhanced by dropout technology. This method, utilizing vibration signals as input, focuses on maintaining recognition ability and quickly adapting to new fault types. BiLSTM, an advanced iteration of LSTM, adeptly captures sequential data information, addressing the impact of distant sequence positions on

the current one and effectively resolving long-term dependency issues. Utilizing an ensemble BiLSTM network enhanced with an attention mechanism, Han et al. [30] devised a robust fault diagnosis method for Variable Refrigerant Flow (VRF) systems, achieving high accuracy and effective generalization under various conditions. Table 1 provides detailed insights into the representative fault diagnosis methods mentioned above, all of which are based on a single network.

2.1.2. From single network to their combinations

A single network turns out to be insufficient when dealing with fault diagnosis issues when they are applied to the real engineering equipment with large categories of data [31,32]: (i) It can hardly represent complex relationships and features appeared in the fault diagnosis processes as different types of problems, sometimes, require different types of networks and a single network is unable to fulfil such as demand; (ii) It processes limited features and requires fuse fusion when it is applied to complicated engineering cases; (iii) It impedes modular design and model function extension. Accordingly, combined networks are proposed to remove the drawbacks introduced by a single network as mentioned-above. Li et al. [33] introduced an approach that combines CNN and GRU for diagnosing gear pitting faults. This method demonstrated impressive fault diagnosis accuracy, reaching 98% even when dealing with limited sample sizes. Wang et al. [34] and Liao et al. [35] extract the local features using CNN, capturing the global features and dynamic information with GRU. Zhang et al. [36] use CNN to extract the collected signal features, and then access LSTM to model the features and get a high fault diagnosis accuracy. You et al. [37] recommended an approach combining CNN, BiLSTM and AM, The robustness and effectiveness of the model was verified by rolling bearing data. Xiang et al. [38] built the CNN-LSTM-AM model to improve the accuracy of fault detection by learning mapping weights and parameters, and to give early warning to abnormal states and detect faults, which was verified on supervisory control and data acquisition (SCADA) data. Wang et al. [39] added WKN to the model to reduce model parameters while dealing with signal noise. Huang et al. [40] carried out adaptive feature refinement of signals through CBAM to realize multi-classification recognition tasks. Xu et al. [41] propose a multi-scale CNN combined with BiLSTM model to solve the problem of insufficient extrapolation ability of intelligent method in fault diagnosis, with improve the generalization performance under complex working conditions. Many scholars have introduced AM into deep learning models to enhance feature expression ability and fault diagnosis accuracy [42,43]. Table 2 lists representative model in the field.

Table 1
Application of single networks in fault diagnosis.

Reference	Model	Object	Signal used	Advantages
Kumar et al. [20]	CNN	Centrifugal pump	Acoustic	Converted input signals into images
Tang et al. [21]	CNN	Hydraulic piston pump	Vibration +Pressure + Sound	Converted input signals into images
Li et al. [22]	CNN	Drilling pump	Vibration	Signal expansion operation
Tang et al. [23]	CNN	Drilling pump	Vibration	Generalized S Transform
Wang et al. [25]	RNN	Wind turbine	Wind speed +Rotor Speed +Generator Speed	Multi-feature input
Bie et al. [27]	LSTM	Reciprocating pump	Vibration	Signal denoising
Miao et al. [29]	GRU	Planetary gearbox	Vibration	Small sample size
Han et al. [30]	BiLSTM	VRF systems.	Thermal Data +Frequency Metrics +Electrical Signals	Enhanced interpretability

2.2. Problem statement and contributions

Drilling pumps are specific devices and have been very a few investigations especially the failure features and fault diagnosis aspects. There is little literature on fault diagnosis of drilling pumps, but its reliable operation is of great significance to drilling engineering, which indicating that intelligent fault diagnosis of drilling pumps is a challenging task but are urgently needed. What's more, signals collected from drilling pumps using sensors in complex and dynamic operating environments inevitably contain noise due to the nature of the pump's operation. This interference can disrupt intelligent fault diagnosis tasks if not reduced. This is because the artificial neural network may interpret noise as a fault-related feature, which could compromise the accuracy of the diagnosis. The reduction of noise in raw signals is crucial for effective drilling pump fault diagnosis. However, all existing noise reduction methods unavoidably increase signal processing time to some extent. This presents a significant challenge for real-time fault diagnosis of drilling pumps, as it reduces the speed of fault identification, which could lead to serious equipment damage or safety incidents. Therefore, there is an urgent need for a fault diagnosis method that can integrate superior noise reduction performance with high efficiency, to aligning with the demands of engineering practice.

The existing state-of-the-art methods, however, encounter in severe limitations when they are applied to the fault diagnosis of drilling pumps, including: (1) Majority of the methods uses vibration signals to characterize the health states. However, drilling pumps pose different failure modes and failure behaviors when compared to other devices. Hence, it is worthwhile to determine the signals that can better reflect their health state; (2) Drilling pumps operate in harsh, uncertain, and unpredictable environments, which challenges of capability of noise reduction of existing methods. To this end, this paper proposes a

Table 2
Application of combined networks in fault diagnosis.

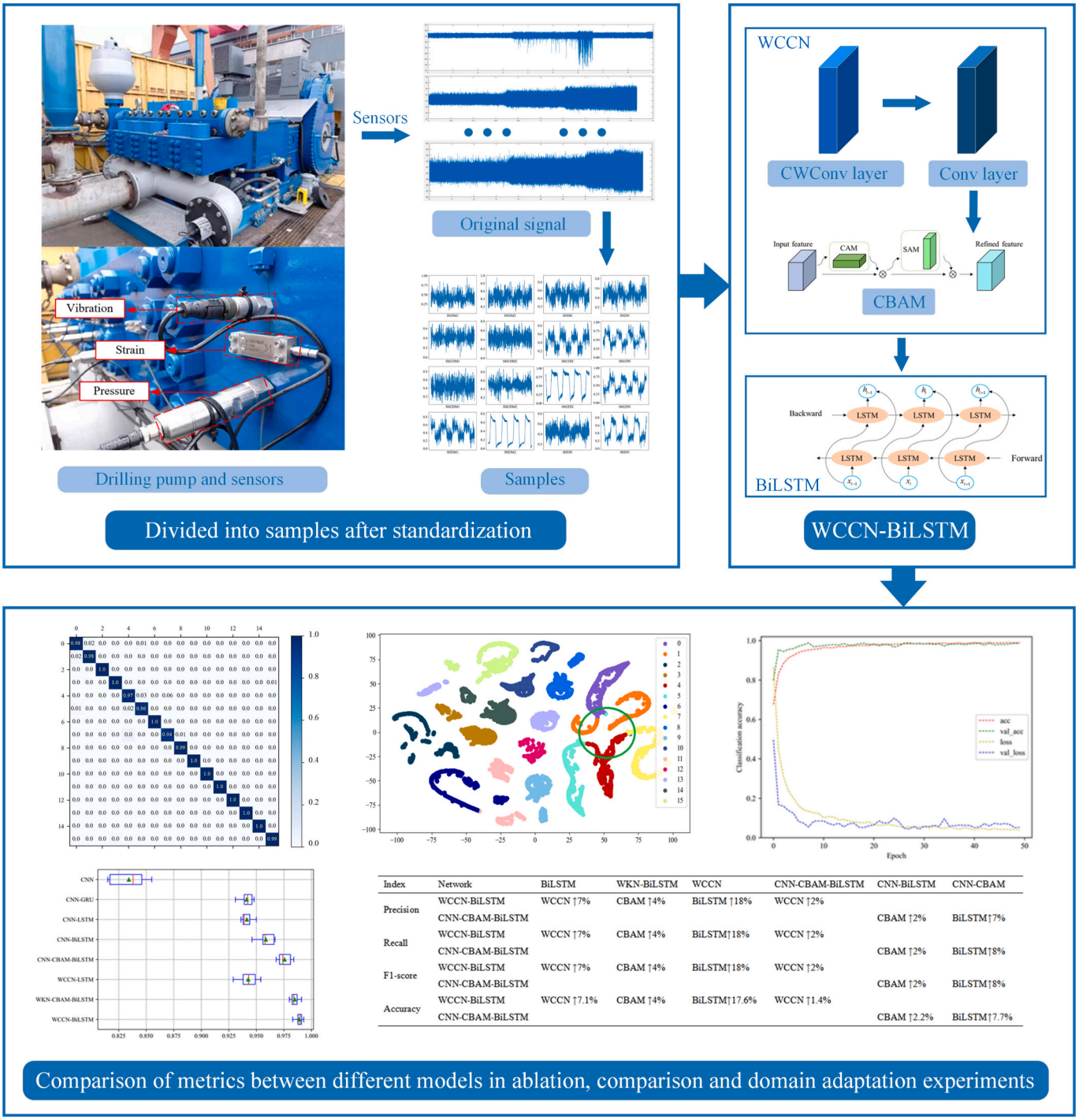
Reference	Model	Object	Signal used	Advantages
Li et al. [33]	CNN-GRU	gear	Acoustic+ vibration	Network composition
Wang et al. [34]	CNN-GRU	chiller system	temperature + pressure	Network composition
Liao et al. [35]	CNN-GRU	hydroelectric generating unit	vibration	Network composition
Zhang et al. [36]	CNN-LSTM	Liquid Rocket Engine	Pressure + power	Simulation data
You et al. [37]	CNN-BiLSTM	Rolling Bearing	vibration	Network composition
Xiang et al. [38]	CNN-LSTM-AM	wind turbine	Temperature + power	Network composition
Wang et al. [39]	WKN-BiLSTM-AM	Rolling Bearing	vibration	Signal denoising
Huang et al. [40]	CNN-CBAM-LSTM	Fiber Vibration Signal Recognition	vibration	Good generalization ability
Xu et al. [41]	MSCNN-BiLSTM	wind turbine bearing	vibration	Good generalization ability
Yang et al. [42]	MBiGRU-AM	bearing	vibration	Introduction of AM
You et al. [43]	BiLSTM-SCN-CBAM	rolling bearing	vibration	Good generalization ability

MDRL: multi-scale deep residual learning; MSCNN: multi-scale convolutional neural network; SCN: sparse convolutional network

WaveletKernelNet-CBAM Net (WCCN) model combining with BiLSTM as a basis of that to build a fault diagnosis framework, namely WCCN-BiLSTM, for drilling pumps. Specifically, the WCCN model incorporates the WKN and CBAM modules. Within this framework, the WKN model seamlessly integrates noise reduction functionality into the network computation. This integration eliminates the need for separate signal processing for noise reduction. Instead, the model autonomously learns and applies the noise reduction function, bypassing additional noise reduction steps and thereby accelerating computation speed. Meanwhile, CBAM enhances the feature representation of the noise-canceled signal in both channel and spatial dimensions, thereby improving the capacity to learn and recognize key features. Following this, the BiLSTM module is employed to further augment the representation of key signal features and conduct sequence modeling, facilitating efficient and accurate fault diagnosis. The contributions of this paper include:

(i) Determine the signal that can best characterizes of drilling pumps' failures based on the Random Forest (RF) model.
(ii) Create a WCCN network combining with the BiLSTM for better noise reduction and feature extraction from time series data.

Overall, this method contributes to safe production of the oil and gas sector and benefit for accurate and robust fault diagnosis of industrial



Index	Network	BiLSTM	WKN-BiLSTM	WCCN	CNN-CBAM-BiLSTM	CNN-BiLSTM	CNN-CBAM
Precision	WCCN-BiLSTM	WCCN ↑7%	CBAM ↑4%	BiLSTM ↑18%	WCCN ↑2%		
	CNN-CBAM-BiLSTM					CBAM ↑2%	BiLSTM↑7%
Recall	WCCN-BiLSTM	WCCN ↑7%	CBAM ↑4%	BiLSTM↑18%	WCCN ↑2%		
	CNN-CBAM-BiLSTM					CBAM ↑2%	BiLSTM↑8%
F1-score	WCCN-BiLSTM	WCCN ↑7%	CBAM ↑4%	BiLSTM↑18%	WCCN ↑2%		
	CNN-CBAM-BiLSTM					CBAM ↑2%	BiLSTM↑8%
Accuracy	WCCN-BiLSTM	WCCN ↑7.1%	CBAM ↑4%	BiLSTM↑17.6%	WCCN ↑1.4%		
	CNN-CBAM-BiLSTM					CBAM ↑2.2%	BiLSTM↑7.7%

Comparison of metrics between different models in ablation, comparison and domain adaptation experiments

Fig. 1. The framework of the fault diagnosis method based on WCCN-BiLSTM.

equipment.

3. Fault diagnosis method based on WCCN-BiLSTM

3.1. WCCN-BiLSTM

In this paper, a new drilling pump fault diagnosis method referred to as WCCN-BiLSTM, was introduced. WCCN employs a continuous wavelet convolutional layer (CWConv) and a standard convolutional layer to reduce noise in the original signal and facilitate basic feature extraction. A Convolutional Block Attention Module (CBAM) is then employed to enhance the feature representation in both channel and spatial dimensions. This enhancement improves the focus on essential features within the data signal while reducing the influence of irrelevant features. Finally, the BiLSTM model is used for the final computation of the feature vectors. After computation, the sequence data is passed through the output layer to derive the signal recognition result. Using Bayesian Optimisation (BO), the model parameters are optimized to achieve optimal model performance. In addition, the incorporation of batch normalization effectively prevents model overfitting. The fault diagnosis framework method based on the WCCN-BiLSTM network is shown in Fig. 1 and with the following steps, see Table 3:

Step 1: Collect the monitoring signals of different degrees of damage to the drilling pump, and select the monitoring signal that can best characterize the drilling pump failure through the RF method.

Step 2: Normalise the monitoring signals and partition the samples, data sets, and labels.

Step 3: Use Bayesian optimization to determine the optimal set of hyperparameters for the WCCN-BiLSTM model.

Step 4: Train and validate the WCCN-BiLSTM model using the compiled drilling pump dataset.

Step 5: Perform ablation experiments to evaluate the effect of each module in the model. Validate the fault diagnosis capability of the model through comparative experiments.

Step 6: Perform domain adaptation experiments to assess the generalization ability of the model across different working condition datasets.

3.2. WaveletKernelNet -CBAM net

CNN has shown success in remaining life prediction and fault diagnosis [44–46]. Typical CNN network comprises an input layer, convolutional layer, pooling layer, fully connected layer, and output layer. The input layer normalizes multidimensional data and feeds the learning data into CNN to guarantee the operational efficiency and learning performance [47,48]. The convolution layer generates features through convolution and using nonlinear activation functions. After convolution, the activation function applies a nonlinear transformation to each convolution's output logarithm, aiming to enhance the linear divisibility of the originally linearly indivisible multidimensional features in another space. CNN is unable to handle noise in raw signals and lacks satisfactory interpretability. To this end, WKN is developed and applied to fault diagnosis [49]. Fig. 2 illustrates that WKN modifies the standard CNN architecture by substituting its first convolutional layer with a Continuous Wavelet Convolutional Layer. This adaptation endows WKN with enhanced noise reduction capabilities. Furthermore, WCCN builds upon the foundation laid by WKN, incorporating a CBAM into the model. This integration enables WCCN to more effectively concentrate on key features.

WCCN adds CWConv layer and CBAM before and after the first convolutional layer compared to CNN, and retains the first convolutional layer and adds CBAM compared to WKN. The CWConv layer includes wavelet kernels with different scales and translation parameters, and performs convolution using a predefined function defined by two learnable parameters. To be specific, the CWConv layer is defined as:

$$h = \psi_{u,s}(t) * x \tag{1}$$

where, $\psi_{u,s}(t)$ represents the predetermined over function to denote results of CWConv layer, the translation and scaling parameters updating of whom is:

$$\begin{aligned} \delta_{u_k} &= \frac{\partial H}{\partial u_k} = \frac{\partial H}{\partial z_k}\frac{\partial z_k}{\partial h_k}\frac{\partial h_k}{\partial \psi^k_{u,s}}\frac{\partial \psi^k_{u,s}}{\partial u_k} \\ \delta_{s_k} &= \frac{\partial H}{\partial s_k} = \frac{\partial H}{\partial z_k}\frac{\partial z_k}{\partial h_k}\frac{\partial h_k}{\partial \psi^k_{u,s}}\frac{\partial \psi^k_{u,s}}{\partial s_k} \end{aligned} \tag{2}$$

$$\begin{cases} u_k = u_k - \eta\delta u_k \\ s_k = s_k - \eta\delta s_k \end{cases} \tag{3}$$

where, ∂ is the derivative operator, $\psi^k_{u,s}$ denotes the k th wavelet kernel of the layer with the length L, S_k and u_k denote the scale and translation parameter, respectively. Specifically, u and s are updated by subtracting the learning rate η and gradientδ.

CBAM is a lightweight general-purpose module to improve representational and generalization capabilities of networks by introducing channel attention and spatial attention [50]. In the WCCN framework, convolutional layers initially process signals, capturing time-dependent features. The channel attention mechanism evaluates each channel via global pooling and a fully connected layer, forming a channel attention graph. This graph is then applied to the original feature maps to selectively enhance or suppress specific channel features. Next, the spatial attention component generates a sequence attention map by performing mean and maximum pooling on the adjusted feature maps. Concatenated along the channel dimension, these results are processed through a convolutional layer. After activation, this map adjusts each temporal point's prominence on the feature map. As shown in Fig. 3.

Within the CBAM framework, input features undergo processing in the channel attention module, high-level features are extracted through average and maximum pooling, and the channel experiences compression and expansion through the multilayer perceptron (MLP). This process enhances the network's capability to represent features across various channels. The features output by MLP are then summed by elements and combined to output feature vectors. The channel attention is computed as:

Table 3
The procedure of the WCCN-BiLSTM network.

Algorithm 1: The proposed method.
Input: WCCN-BiLSTM, with best hyperparameters selected by BO; Dataset of drilling pump;
Data Preprocessing: Monitoring Signal Selection; Signal Standardization;
for epoch = 1, 2, …, epoch **do:**
Data undergoes noise reduction and feature extraction through WCCN channel, followed by feature enhancement;
Feature vector operations are performed on data output from the WCCN channel using the BiLSTM channel;
Obtaining output of WCCN-BiLSTM;
Adam algorithm is applied to optimize and reduce training error;
end
Output: Fault diagnosis of drilling pumps

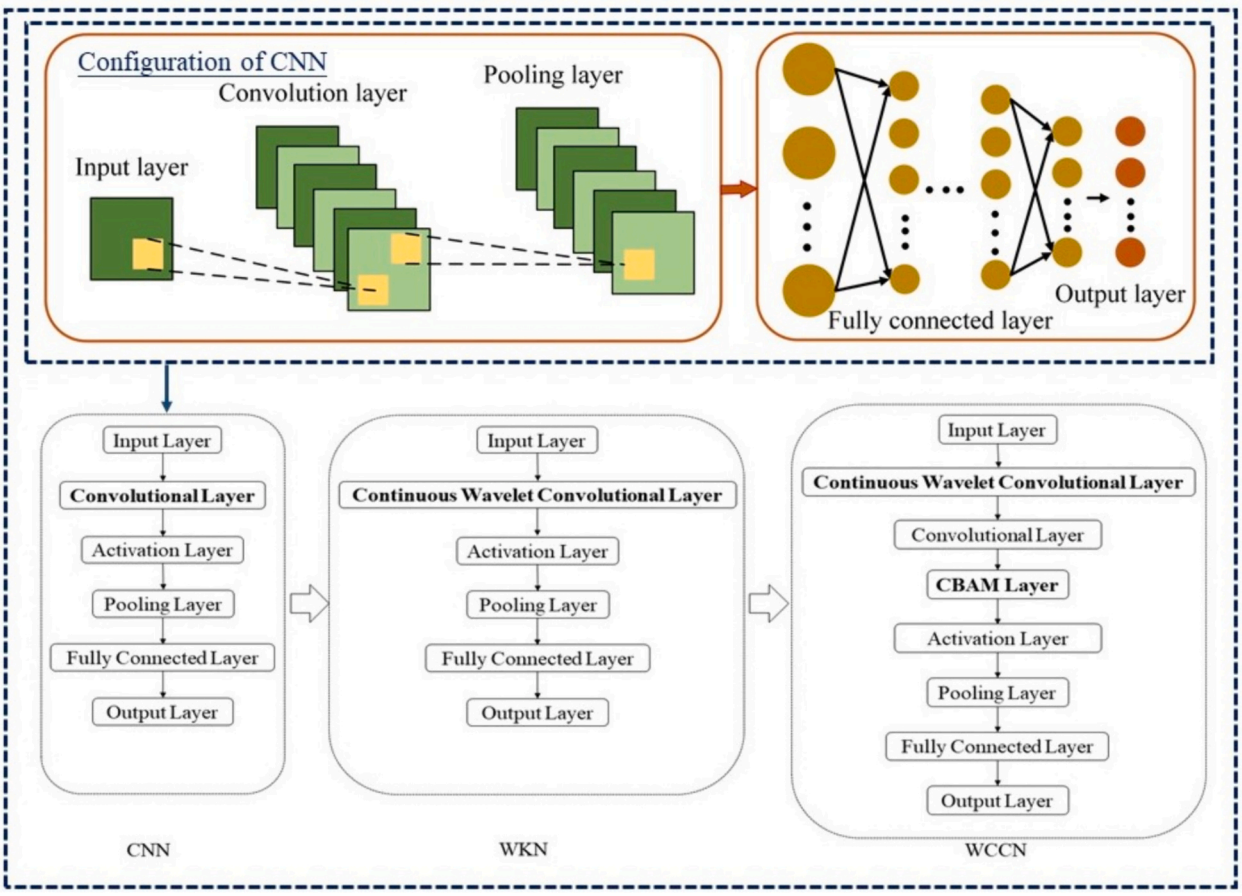


Fig. 2. Developments from CNN to WKN and WCCN.

$$M_c(F) = \sigma(MLP(AvgPool(F))) + MLP(MaxPool(F)) = \sigma\left(W_1\left(W_0\left(F^c_{avg}\right)\right)\right) + W_1\left(W_0\left(F^c_{max}\right)\right) \quad (4)$$

where, $M_c(F)$ is the channel attention weight, σ denotes the sigmoid activation function, and F reflects the feature matrix.

The spatial attention module explores the intrinsic relationship of features, which is computed by:

$$M_s(F) = \sigma(Conv([AvgPool(F); MaxPool(F)])) = \sigma\left(Conv\left(\left[F^s_{avg}; F^s_{max}\right]\right)\right) \quad (5)$$

where, $M_c(F)$ is the spatial attention weight, and $Conv(\cdot)$ displays the convolution operation.

3.3. BiLSTM

LSTM is capable to process and model time series data with especially in solving long-term dependency problems [51–53], composes of different gated units, see Fig. 4(a). The input gate determines new input information into the cell state; The forgetting gate ascertains retains or not the previous information; The output gate determines outputs to the network layer in the time slice. These mechanisms enable LSTMs managing previous information and update the cell state if needed. The calculation of LSTM are as follows:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \quad (6)$$

where f_t and x_t represent output and input. σ reflects gate pass function, h_{t-1} is the output. W_f and b_f is the weight and bias matrix of forget gate.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (7)$$

where i_t denotes the information input , W_i and b_i is the weight and bias matrix of input gate.

$$\widetilde{C}_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (8)$$

where $\widetilde{C}_t$ is candidate memory cell. W_c and b_c is the weight and bias matrix of input gate.

$$C_t = f_t \times C_{t-1} + i_t \times \widetilde{C}_t \quad (9)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (10)$$

where o_t is the information output, C_t represents the current state of the cell.

$$h_t = o_t \times tanh(C_t) \quad (11)$$

where h_t represents the ultimate data derived from both o_t and C_t.

BiLSTM introduces a hidden layer in two directions on the basis of LSTM to better capture the before and after relationships in time series data. The BiLSTM structure is shown in Fig. 4(b).

4. Data processing

4.1. Drilling pump & data

The experimental data of this paper was experimented and collected from the HH2400 drilling pump test platform. This dataset was obtained from drilling components returned for repair and testing to replicate

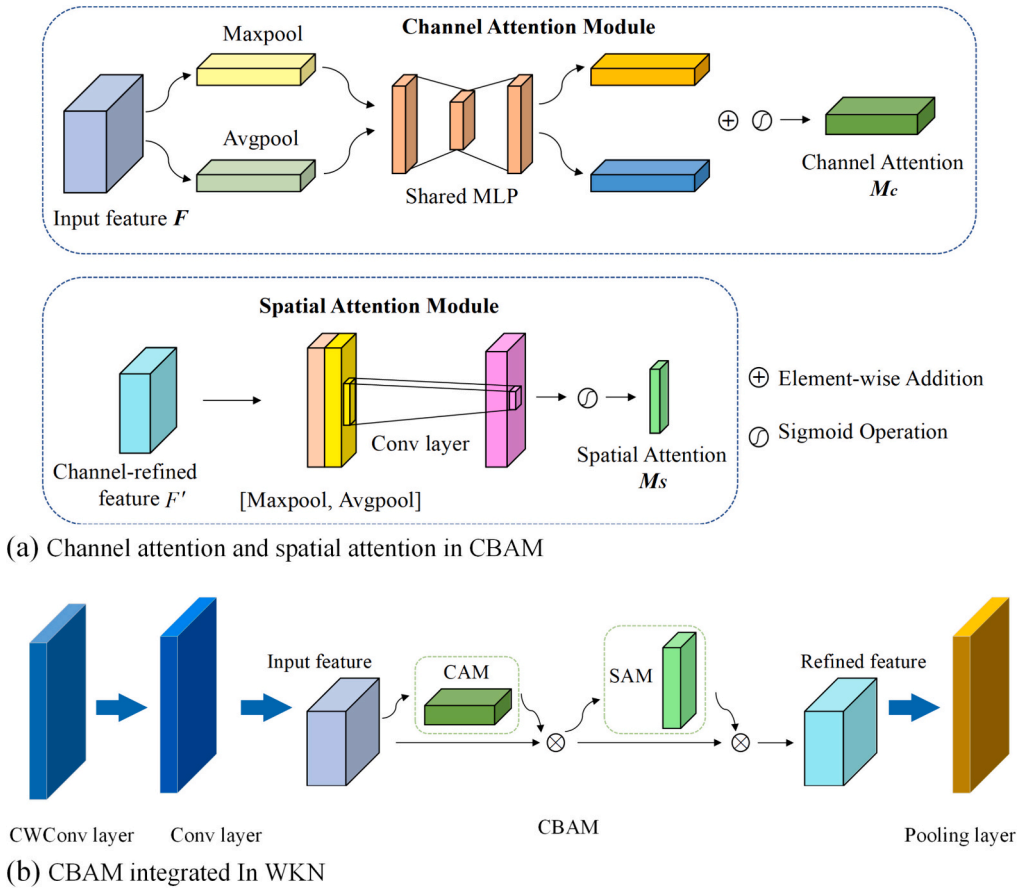


Fig. 3. Frameworks of CBMA and WCCN.

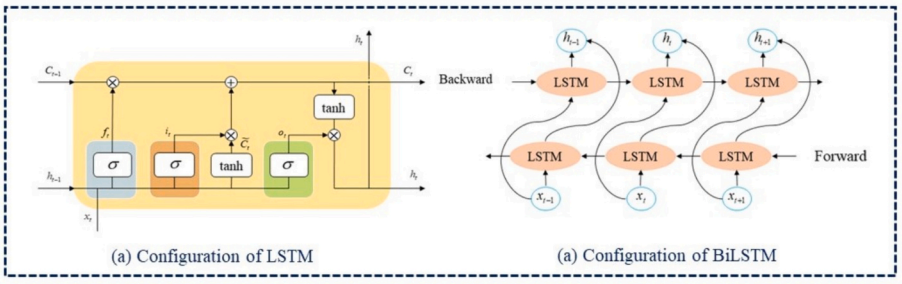


Fig. 4. Structural of LSTM and BiLSTM.

actual operating conditions. The Siemens 1500 series PLC is employed for achieving real-time data sampling with a period of 1ms or shorter. The data acquisition module is equipped with pressure sensors, strain sensors, vibration sensors, etc. on each cylinder. The HH2400 drilling pump and its sensors configuration are shown in Fig. 5(a). For simulation signals, a 16-bit precision extended analog input module is utilized, featuring 8 channels per module and a total of 3 modules, ensuring data accuracy. Following initial processing, the sampled data is stored in a dedicated data storage block (DB) within the PLC for communication transit, facilitating subsequent retrieval by the host computer. Utilizing Ethernet as the carrier, OPC is employed to efficiently coordinate the rapid retrieval of PLC data, with a requirement for an effective data retrieval cycle of within 10ms. Field engineers establish data analysis rules and algorithms to determine the status of the cylinder valve based on data anomalies, classifying them into categories such as normal operation, mild damage, moderate damage, and serious damage.

Monitoring data obtained from nine groups of drilling pumps, each operating under four different conditions. A pump stroke is subjected to different frequencies: 50 strokes per minute (SPM), 90 SPM, 110 SPM, and 130 SPM. Valve states include normal, minor damage, moderate damage, and heavy damage, see Fig. 5(b). The 16 failures of the drilling pump are listed in Table 4.

In actual drilling operations, 70–130 SPM pumping stroke is the most common working situation. Accordingly, dataset C, with 110SPM, was selected to verify the diagnostic capability of the WCCN-BiLSTM model. Each operating condition of drilling pump contains 16 states with 160,000 samples and a single sample contains 2048 data sampling points. The total sample size was partitioned into training, validation, and test sets using an 8:1:1 ratio, resulting in 128,000, 16,000, and 16,000 samples, respectively.

Fig. 5. Sensor configuration and failures of the HH2400 drilling pump/ (a) Sensor configuration; (b) Failures.

Table 4
Partition of drilling pump datasets.

Dataset	Pump stroke (SPM)	State		Label	Encoder
		Suction valve	Discharge valve		
A/B/C/D	50/90/110/130	Normal	Minor damage	SNDM1	0
		Normal	Moderate damage	SNDM2	1
		Normal	Heavy damage	SNDH	2
		Normal	Normal	SNDN	3
		Minor damage	Minor damage	SM1DM1	4
		Minor damage	Moderate damage	SM1DM2	5
		Minor damage	Heavy damage	SM1DH	6
		Minor damage	Normal	SM1DN	7
		Moderate damage	Minor damage	SM2DM1	8
		Moderate damage	Moderate damage	SM2DM2	9
		Moderate damage	Heavy damage	SM2DH	10
		Moderate damage	Normal	SM2DN	11
		Heavy damage	Minor damage	SHDM1	12
		Heavy damage	Moderate damage	SHDM2	13
		Heavy damage	Heavy damage	SHDH	14
		Heavy damage	Normal	SHDN	15

S: Suction valve; D: Discharge valve; N: Normal; M1: Minor damage; M2: Moderate damage; H: Heavy damage

4.2. Monitoring signals selection

The study acquires vibration, pressure, strain, crankshaft angle, and relieve pressure signals for drilling pumps. The random forest (RF) [54] method is utilized to identify the monitoring signals that can best reflect the fault characteristics of the drilling pump. A RF consists of several unrelated decision trees built randomly. Each decision tree in the forest judges a new sample separately to determine its class, and the final classification result is obtained by voting. In regression problems, the RF calculates the average of all decision tree outputs, quantifying the importance of the features. The higher the feature importance, the more significant the contribution to the accurate prediction of the model, and vice versa. The steps for building RF are as follows:

Step 1: Utilize random sampling method to select bootstrap samples to construct decision tree.

Step 2: Split decision trees until all samples in a node belong to the same class.

Step 3: Classify samples of the RF using the "with-replacement" method.

The Gini index used to evaluate performance of classification and regression is defined by:

$$\text{Gini} = 1\text{-}\sum P_i^2 \tag{12}$$

where, P_i represents the proportion of the *i-th* class samples in the dataset at the current node.

In RF classification, the average of all Gini index weights the indicator, by:

$$Q_k = \frac{\sum_{i}^{n}\sum_{j=1}^{i} D_{Gkij}}{\sum_{k}^{m}\sum_{i}^{n}\sum_{j=1}^{i} D_{Gkij}} \tag{13}$$

where, Q_k represents the importance of the *k-th* indicator among the all, *m* represents the total number of indicators, which is 5 in this paper, *n* reflects the number of decision trees, *t* is the nodes of each classification tree. D_{Gkij} represents the decrease in Gini index of the *k-th* indicator at the *j-th* node of the *i-th* tree.

Fig. 6 displays the importance of signals for drilling pumps' fault diagnosis. The findings indicate that the strain signal is the most effective in capturing the fault characteristics of drilling pumps. Selecting the most relevant and suitable sensor to collect information for fault diagnosis can simplify the modeling of fault diagnosis models and reduce the

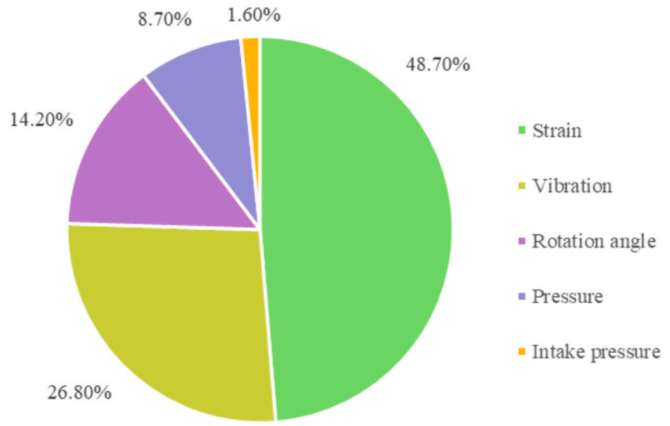


Fig. 6. The importance of signals for drilling pumps' fault diagnosis.

complexity in data processing and feature extraction and contribute to real-time monitoring and rapid diagnosis. Thus, strain signals are used for drilling pump fault diagnosis in this study.

5. Experimental section

5.1. Bayesian optimization model hyperparameters

Bayesian optimization (BO) is widely recognized as a highly effective approach to hyperparameter selection in machine learning models [55]. Its notable strengths include the ability to achieve objectives with fewer iterations, ensuring faster convergence. In addition, BO is particularly robust to non-convex challenges, making it a preferred choice in complex optimization scenarios. Furthermore, a distinct advantage of BO lies in its use of a priori information about the parameters. This process involves continuously updating the posterior distribution of the objective function by adding new samples, which helps to closely match the true distribution. Consequently, in this paper, BO is used as a method for selecting model hyperparameters to find the most optimal set of hyperparameters for the model.

In this study, the BO process was structured in two distinct stages. Initially, five initial optimization iterations were conducted to establish a baseline model of parameter performance. This phase is essential, as Bayesian optimization relies on prior data to commence the iterative procedure. Following this, a further 50 optimization iterations were executed. These iterations were systematically designed by the Bayesian optimization algorithm, progressively enhancing and refining the parameter choices to more accurately approach the optimal solution. Altogether, 55 iterations were completed, a number deemed appropriate for balancing efficient exploration of the parameter space against computational costs. The main hyperparameters of the WCCN-BiLSTM model are selected by BO, including the filters, kernel size, strides, and units of BiLSTM. Fig. 7 represents the process by which the accuracy of the validation set corresponding to each set of hyperparameters selected by the network model using the Bayesian algorithm in the hyperparameter space changes. This figure comprises two sections: the lower section details the hyperparameter values, while the upper section highlights the peak accuracy achieved with each hyperparameter set during validation. The model, in each iteration, BO dynamically adjusts the value of the next hyperparameter selection based on the previously selected hyperparameters with their achieved highest accuracy. Table 5 comprehensively presents the range of hyperparameter values explored and the conclusive outcomes of this process.

5.2. Model evaluation method

The model's convergence was assessed using the cross-entropy loss function:

Table 5
Hyperparameter ranges in BO and the optimized results [56].

No.	Name	Range	Results
1	filters	[1,64]	28
2	kernel size	[1,64]	33
3	strides	[1, 16]	12
4	LSTM	[1,64]	36

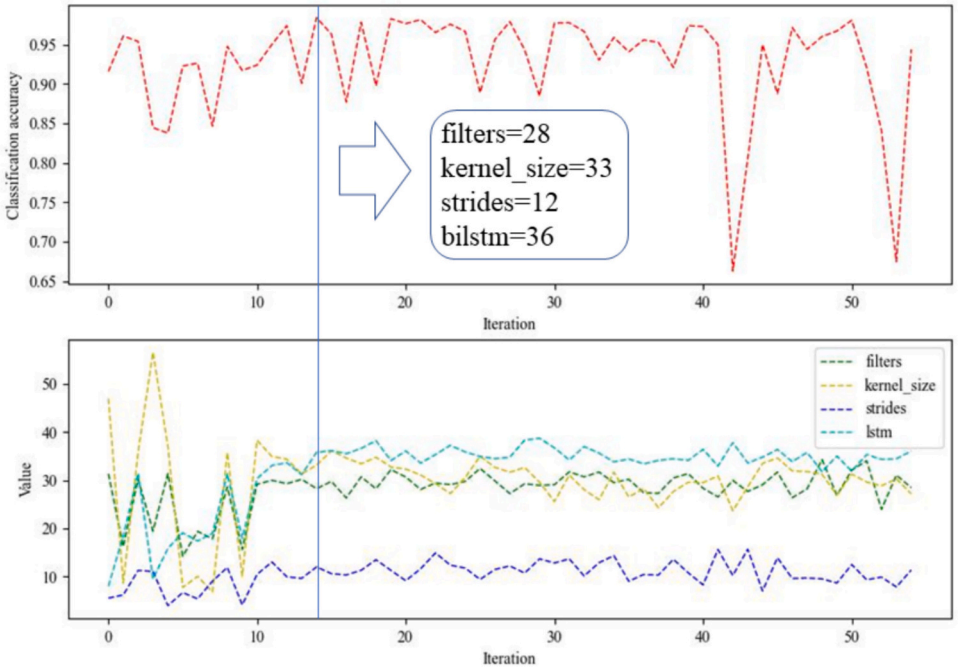


Fig. 7. The hyperparameter training process.

$$loss = -\sum_{i=1}^{K} y_i log(p_i) \tag{18}$$

where P_i is the probability value of the label i, while the label is i, they$_i$is 1, otherwise is 0.

Assess the model with the following criteria: *accuracy*, *precision*, *recall* and $F1-score$.

$$accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{19}$$

$$precision = \frac{TP}{TP+FP} \tag{20}$$

$$recall = \frac{TP}{TP+FN} \tag{21}$$

$$F1-score = \frac{2^{*}precision}{precision+recall} \tag{22}$$

where TP, FP, FN, TN denote the counts of true positive, false positive, false negative, and true negative results, respectively.

In Eq. (19), accuracy signifies the ratio of correctly predicted samples to the total number of samples. Eq. (20) defines precision, representing the proportion of correctly predicted positive samples to the actual predicted positive samples. Eq. (21) defines recall as the proportion of positive samples that are correctly predicted, and Eq. (22) formulates the F1-Score as a comprehensive performance metric for the model.

5.3. Ablation experiment

To verify the structure of the model, ablation experiments were designed to demonstrate the active role of each individual module. The experiment is centered around Dataset C, corresponding to a pump stroke of 110SPM. The training process for the WCCN-BiLSTM model is depicted in Fig. 8, where the accuracy and loss values are presented. As indicated by the figure, a noticeable convergence trend is evident once the iteration count reaches 50.

Ablation experiments for the WCCN, CBAM, and BiLSTM modules were conducted on Computer 1, aiming to evaluate the individual effectiveness of each module in the developed WCCN-BiLSTM model. Computer 1 was outfitted with an i5-12400F processor, 32GB of RAM, and a 3060 GPU. Table 6 presents a quantitative comparison of *precision*, *recall*, and *f1-score* values across various models using Dataset C, demonstrating that the WCCN, CBAM, and BiLSTM modules enhanced fault diagnosis accuracy by 7.1%, 4%, and 17.6%, respectively. For more comprehensive comparisons, refer to Table 7.

5.4. Comparison experiment

5.4.1. Comparison of other networks

In this section, the performance of the WCCN-BiLSTM model is evaluated by comparing it with seven other models. Fig. 9(a) demonstrates that the proposed model, across different configurations, achieves faster convergence in accuracy compared to other models. Additionally, it exhibits notable stability, showing minimal fluctuation in accuracy post-convergence, unlike other methods. This suggests superior convergence and robustness of the proposed method. The following experiments and analyses focus on illustrating the performance of each model using dataset C, which represents common working conditions. Fig. 9(b) graphically presents the average precision, recall, and F1 score for both the state-of-the-art and the proposed models. Table 8 offers a thorough comparative analysis of time consumption for different models, evaluating their computational efficiency. To gauge real-world performance, the test set was run on two computers with distinct specifications. Computer 1 featured an i5-12400F processor, 32GB of RAM, and a 3060 GPU, whereas Computer 2 was equipped with an i9-13900KF processor, 48GB of RAM, and a 4090 GPU.

To mitigate the impact of randomness from data sampling order and initial weight settings, which can cause variation in results across experiments, conducting multiple experiments and calculating their average can effectively reduce this variability. Therefore, under identical parameter settings, five independent experiments were conducted, and the results of the fault diagnosis for the test set in Dataset C are displayed in Fig. 9(c). The figure reveals that the WCCN-BiLSTM model exhibits the greatest stability and highest average accuracy at 98.9%. In

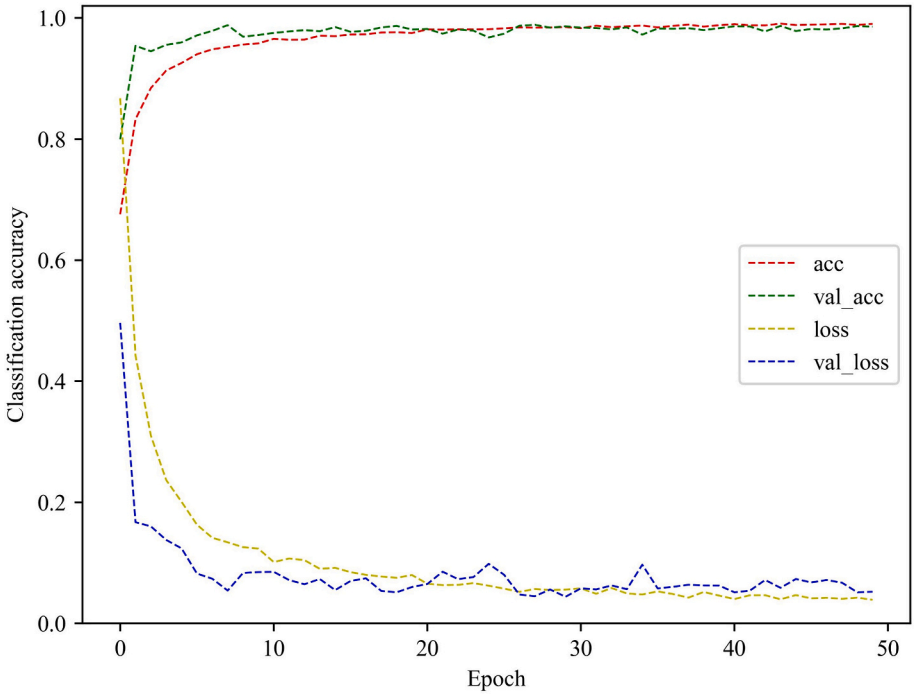


Fig. 8. The variation of values throughout the model training.

Table 6
The quantitative comparison results of each module ablation experiment.

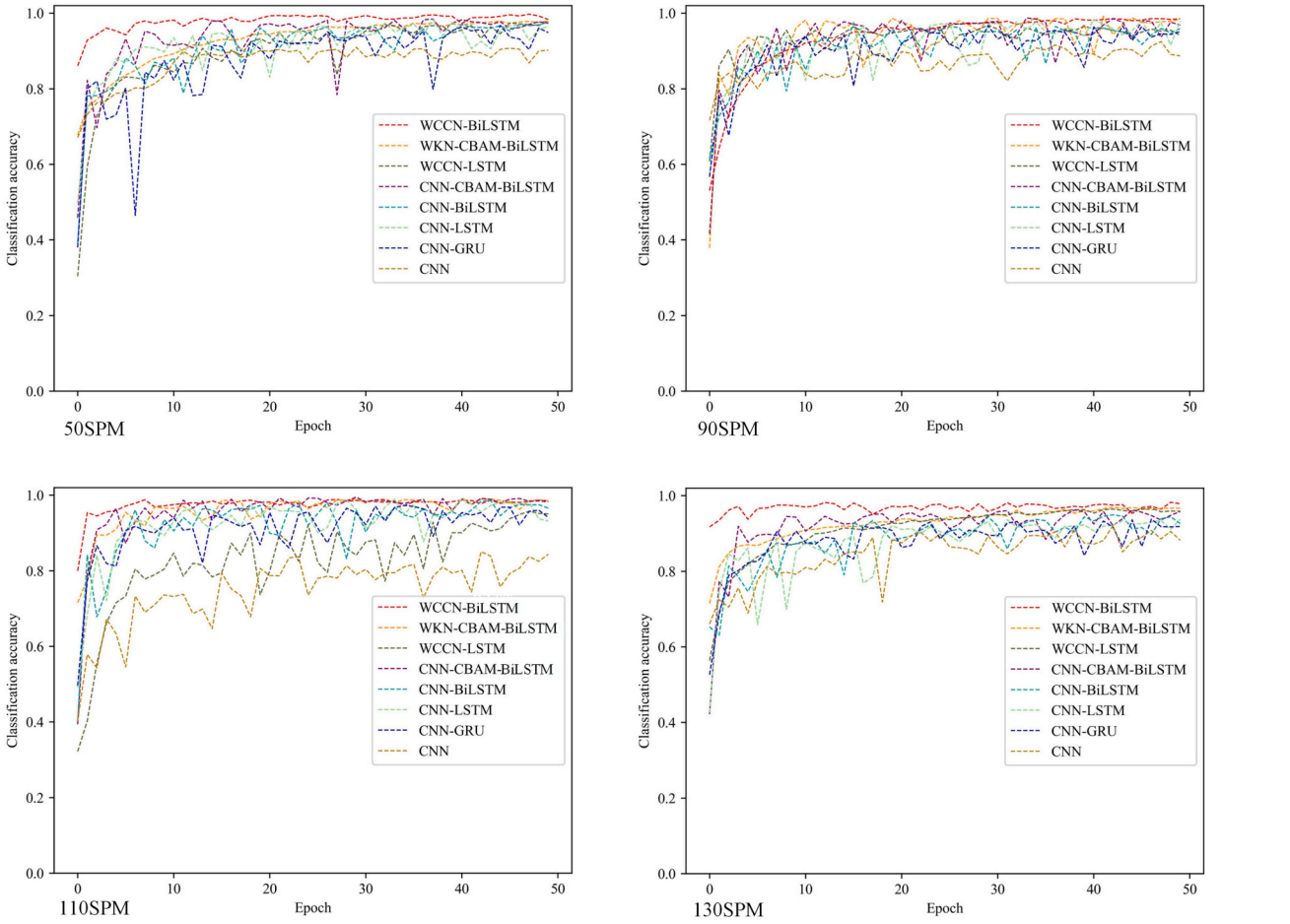
Index	Network	BiLSTM	WKN-BiLSTM	WCCN	CNN-CBAM-BiLSTM	CNN-BiLSTM	CNN-CBAM
Precision	WCCN-BiLSTM CNN-CBAM-BiLSTM	WCCN ↑7%	CBAM ↑4%	BiLSTM ↑18%	WCCN ↑2%	CBAM ↑2%	BiLSTM↑7%
Recall	WCCN-BiLSTM CNN-CBAM-BiLSTM	WCCN ↑7%	CBAM ↑4%	BiLSTM↑18%	WCCN ↑2%	CBAM ↑2%	BiLSTM↑8%
F1-score	WCCN-BiLSTM CNN-CBAM-BiLSTM	WCCN ↑7%	CBAM ↑4%	BiLSTM↑18%	WCCN ↑2%	CBAM ↑2%	BiLSTM↑8%
Accuracy	WCCN-BiLSTM CNN-CBAM-BiLSTM	WCCN ↑7.1%	CBAM ↑4%	BiLSTM↑17.6%	WCCN ↑1.4%	CBAM ↑2.2%	BiLSTM↑7.7%

Table 7
Quantitative evaluation results of ablation experiments were performed on dataset C.

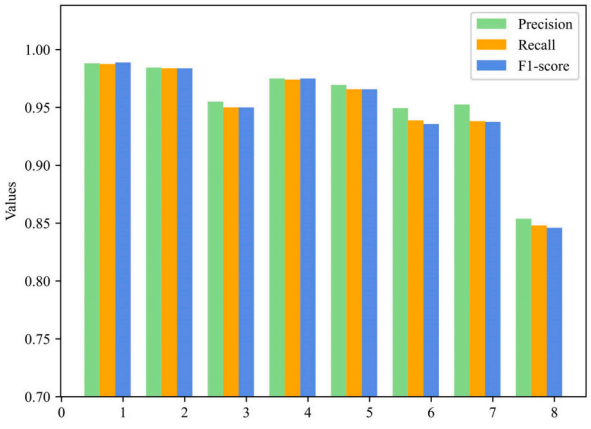
Network	Structure	Input	Output	Precision	Recall	F1-score	Accuracy
WCCN-BiLSTM	**CWConv (filters = 32, kernel size = 64, stride = 1)** **Conv (filters = 28, kernel size = 33, stride = 12) CBAM ()** **Pooling (pool size = 2, strides = 2)** **BiLSTM (units = 32)** **Dense ()**	**(2048, 1)** **(2048, 32)** **(171, 28)** **(171, 28)** **(85, 28)** **(85, 72)**	**(2048, 32)** **(171, 28)** **(171, 28)** **(85, 28)** **(85, 28)** **(6120, 16)**	**0.99**	**0.99**	**0.99**	**0.988**
BiLSTM	Pooling (pool size = 2, strides = 2) BiLSTM (units = 32) Dense ()	(2048, 1) (1024, 1) (1024, 72)	(1024, 1) (1024, 72) (73,728, 16)	0.92	0.92	0.92	0.917
WKN-BiLSTM	CWConv (filters = 32, kernel size = 64, stride = 1) Conv (filters = 28, kernel size = 33, stride = 12) Pooling (pool size = 2, strides = 2) BiLSTM (units = 32) Dense ()	(2048, 1) (2048, 32) (171, 28) (85, 28) (85, 72)	(2048, 32) (171, 28) (85, 28) (85, 28) (6120, 16)	0.95	0.95	0.95	0.948
WCCN	CWConv (filters = 32, kernel size = 64, stride = 1) Conv (filters = 28, kernel size = 33, stride = 12) CBAM () Pooling (pool size = 2, strides = 2) Dense ()	(2048, 1) (2048, 32) (171, 28) (171, 28) (85, 28)	(2048, 32) (171, 28) (171, 28) (85, 28) (2380, 16)	0.81	0.81	0.81	0.812
CNN-CBAM-BiLSTM	Conv (filters = 32, kernel size = 64, stride = 1) Conv (filters = 28, kernel size = 33, stride = 12) CBAM () Pooling (pool size = 2, strides = 2) BiLSTM (units = 32) Dense ()	(2048, 1) (2048, 32) (171, 28) (171, 28) (85, 28) (85, 72)	(2048, 32) (171, 28) (171, 28) (85, 28) (85, 28) (6120, 16)	0.97	0.97	0.97	0.974
CNN-BiLSTM	Conv (filters = 32, kernel size = 64, stride = 1) Conv (filters = 28, kernel size = 33, stride = 12) Pooling (pool size = 2, strides = 2) BiLSTM (units = 32) Dense ()	(2048, 1) (2048, 32) (171, 28) (85, 28) (85, 72)	(2048, 32) (171, 28) (85, 28) (85, 28) (6120, 16)	0.97	0.97	0.97	0.966
CNN-CBAM	Conv (filters = 32, kernel size = 64, stride = 1) Conv (filters = 28, kernel size = 33, stride = 12) CBAM () Pooling (pool size = 2, strides = 2) Dense ()	(2048, 1) (2048, 32) (171, 28) (171, 28) (85, 28)	(2048, 32) (171, 28) (171, 28) (85, 28) (2380, 16)	0.92	0.91	0.91	0.911

contrast, the average accuracies of the other seven models are 83.4%, 94.1%, 94.1%, 95.9%, 97.6%, 94.3%, and 98.5%, respectively.

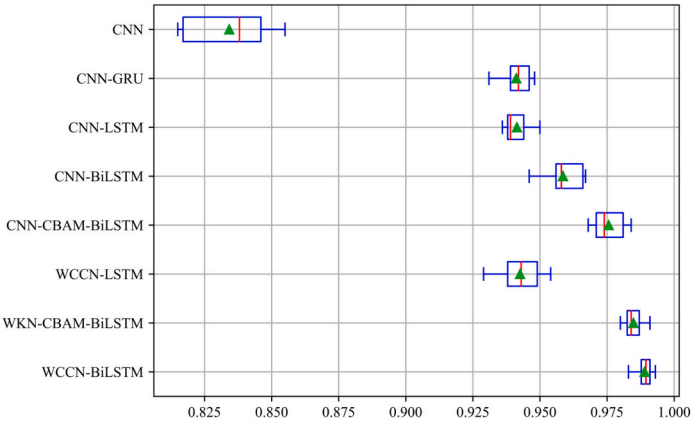
The confusion matrix, a tool for comparing actual versus predicted classifications, aids in visualizing and assessing classification accuracy. To gauge the performance of various models, we present the confusion matrix for dataset C (110SPM) under relevant working conditions in Fig. 10. This figure highlights the classification accuracy for each label class, with a special focus on those falling below 95% accuracy. It shows that multiple labels in other models do not meet the 95% accuracy mark. However, the proposed method has just one label with a 94% accuracy, slightly below 95%. These results underscore the WCCN-BiLSTM model's superior classification capabilities compared to other methods.



(a) Accuracy values with different models



(b) Values of different evaluation indexes of models in 110SPM



(c) fault diagnosis results/ The red/green line: median/average in 110SPM

Fig. 9. Performance indicators of various models in the comparison experiment.

A visualization technique, the t-SNE algorithm, introduces a shift from the high-dimensional Euclidean distance among data points to the conditional probability of their similarity [57]. It is effective in mapping high-dimensional data to low-dimensional data while preserving the relative similarity of the original data. To assess the feature extraction capability of the proposed method, we utilized the t-SNE dimensionality reduction technique, as depicted in Fig. 11. This figure showcases the varying classification abilities of different models with respect to data labeled differently. The visualization distinctly demonstrates how some models, like CNN, have a lesser ability to differentiate between labels, leading to a clustering of varied labels. In contrast, the proposed method is significantly more adept in T-SNE visualization, achieving clear clustering of the same labels and reducing overlap between different labels, thereby showcasing its efficacy in drilling pump fault diagnosis. Fig. 11 also reveals that the WCCN-BiLSTM model has certain limitations in classifying data corresponding to label 4, showing overlap with labels 0 and 7, while other models also struggle to differentiate effectively. Despite this, the WCCN-BiLSTM model generally succeeds in classifying and identifying different labels effectively, unlike the seven comparison models which demonstrate varying degrees of overlap.

Table 8
Comparison of quantitative evaluation results of different models.

Network	Precision	Recall	F1-score	Time1	Time2
WCCN-BiLSTM	**0.9881**	**0.9875**	**0.9888**	**5.98 s**	**4.68 s**
WKN-CBAM-BiLSTM	0.9844	0.9838	0.9838	5.76 s	4.58 s
WCCN-LSTM	0.9550	0.9500	0.9500	5.05 s	3.92 s
CNN-CBAM-BiLSTM	0.9750	0.9740	0.9750	5.89 s	4.78 s
CNN-BiLSTM	0.9694	0.9656	0.9656	5.36 s	4.06 s
CNN-LSTM	0.9494	0.9388	0.9356	4.63 s	3.17 s
CNN-GRU	0.9525	0.9381	0.9375	4.43 s	3.01 s
CNN	0.8538	0.8480	0.8460	3.85 s	2.34 s

Overall, this confirms that the proposed method achieves denser clustering in feature space, which translates to more accurate fault identification in drilling pumps.

5.4.2. Comparison of other denoising methods

To comprehensively evaluate the noise reduction effect of the WCCN-BiLSTM model, a denoising comparison experiment was designed. Initially, the original signals were processed individually using different noise reduction techniques. Subsequently, the denoised signals from each method were inputted into the CNN- BiLSTM model to generate their respective outputs. Following this, a comparison was made between these output results and the processing outcomes of the WCCN-BiLSTM model on the original signal. Through comparing and analyzing the fault diagnosis effects under different noise reduction methods, a clearer judgment can be made on the noise reduction effect, enabling a more comprehensive evaluation of the noise reduction performance of the WCCN-BiLSTM model. Table 9 delineates the specifics of the comparison among different methods. It also indicates that the proposed method performs favorably compared to other noise reduction techniques, exhibiting no lag in any of the metrics. Additionally, it

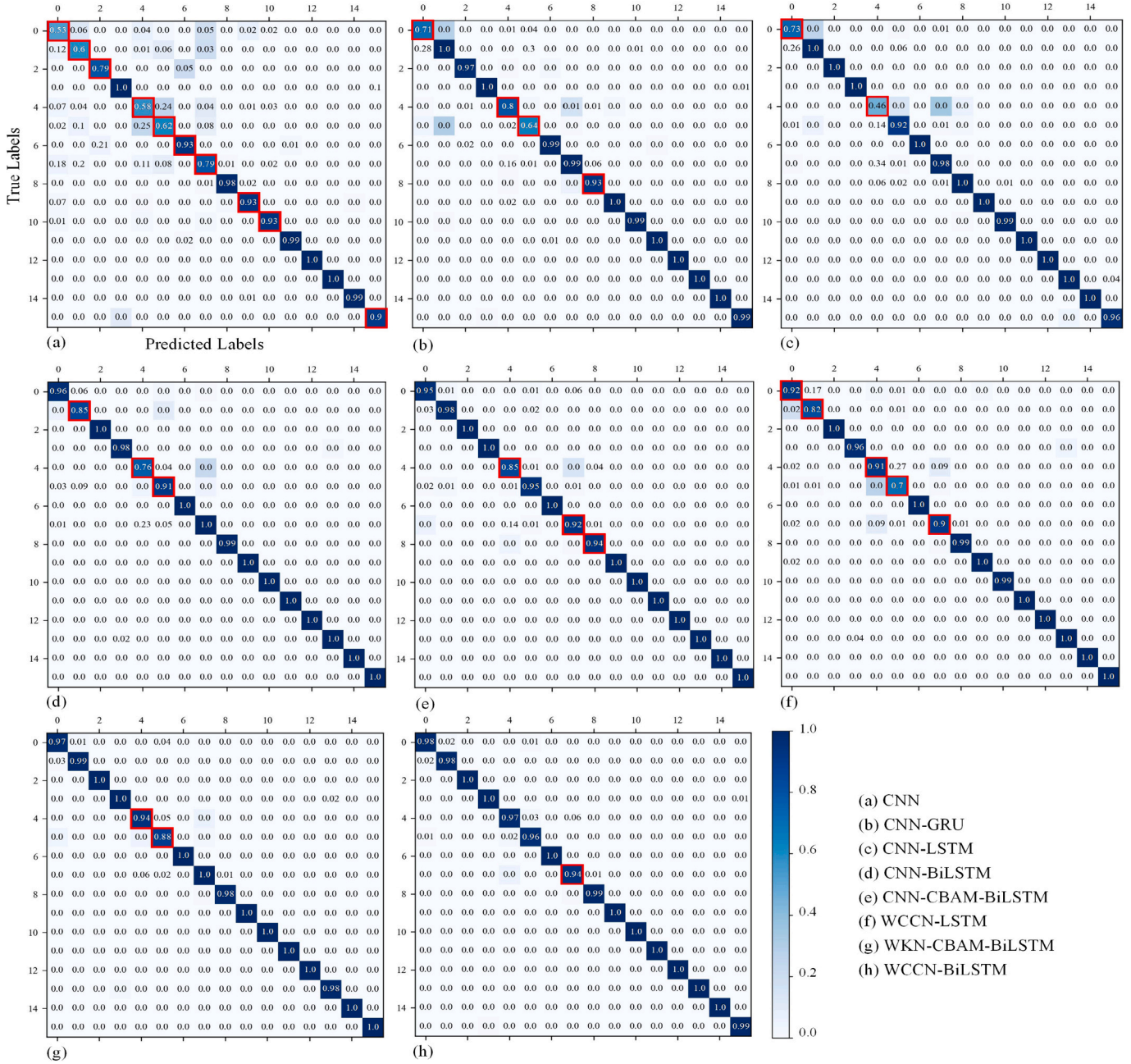


Fig. 10. Confusion matrix of models.

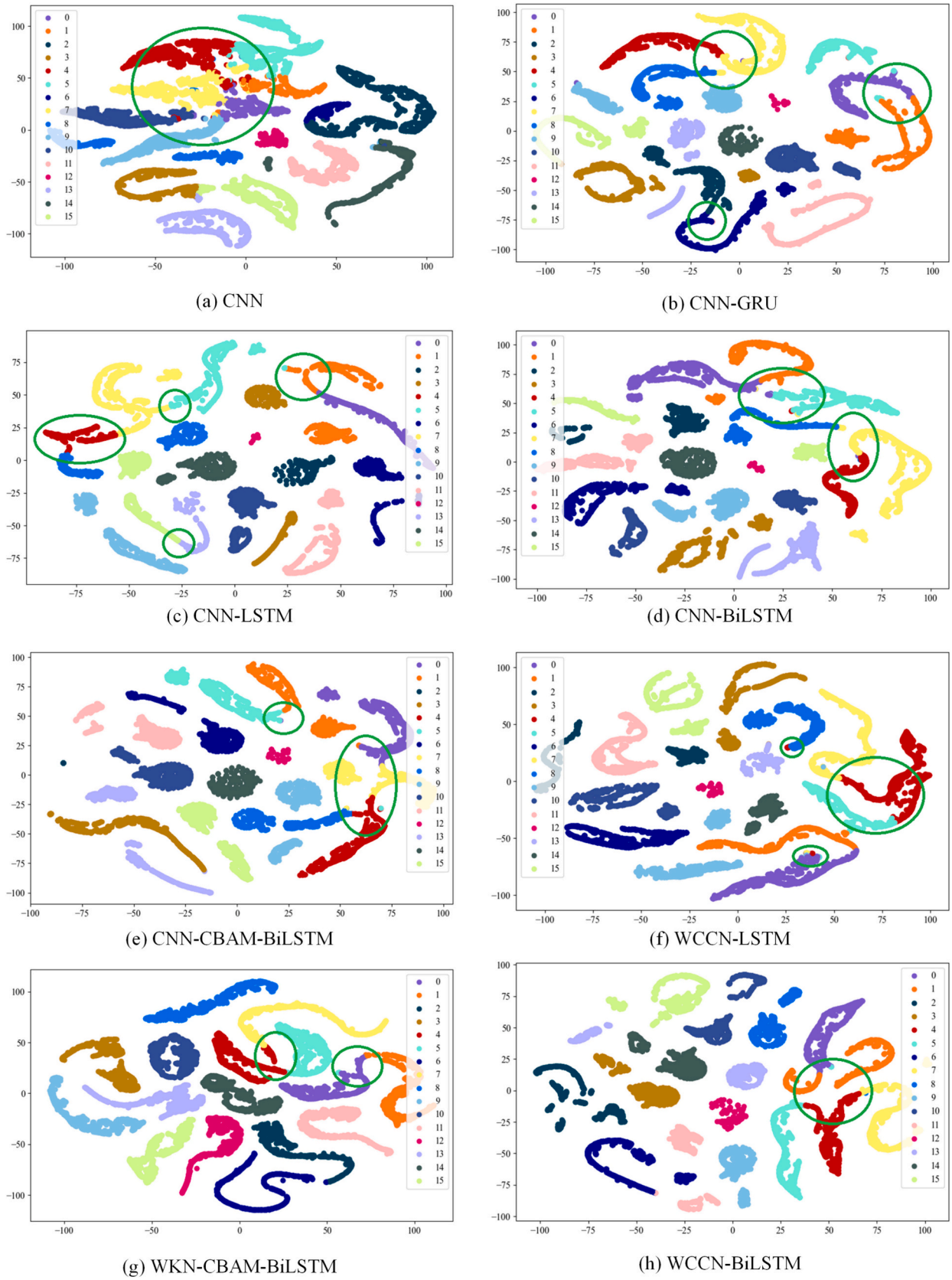


Fig. 11. Comparison of fault identification.

Table 9
Comparing of various noise reduction techniques.

Method	Accuracy	Precision	Recall	F1-score	Time1	Time2
WCCN-BiLSTM	**0.9888**	**0.9892**	**0.9888**	**0.9888**	**5.98 s**	**4.68 s**
FFT-CNN-BiLSTM	0.9849	0.9852	0.9849	0.9849	9.45 s	7.45 s
EMD-CNN-BiLSTM	0.9888	0.9891	0.9888	0.9888	148.7 s	51.36 s
WT-CNN-BiLSTM	0.9879	0.9882	0.9879	0.9878	9.25 s	6.07 s
CNN-BiLSTM	0.9656	0.9694	0.9656	0.9656	5.25 s	3.86 s

becomes evident that direct fault diagnosis of the raw signal from the drilling pump without employing noise reduction techniques is less effective. These findings underscore the effectiveness of the proposed method in signal noise reduction, which bears positive significance for fault diagnosis in drilling pumps.

Given the operational conditions in the field for drilling pumps, signal preprocessing can extend fault diagnosis time, leading to delays and potentially serious accidents. Therefore, this subsection evaluates not only the accuracy and other metrics of each noise reduction method but also conducts a comparative analysis of the time-consuming performance of these methods on two different computers. The computers utilized are the same as those described in Section 5.3.1. Specifically, Table 9 documents the average time required by each method to process the test set data from reading to completing recognition, providing valuable insights into the effectiveness of these approaches. The figure illustrates that upon applying the noise reduction technique, the running time of the different methods on the test set increases to varying degrees. In particular, the EMD denoising technique requires significantly more time on computers 1 and 2, 13 and 28 times longer, respectively—compared to performing without noise reduction. This substantial increase in time expenditure notably elevates the overall time cost of drilling pump fault diagnosis. In comparison to other noise reduction techniques, the built-in noise reduction method of WCCN-BiLSTM exhibits minimal impact on signal processing consumption time, demonstrating consistent computational efficiency across the comparison. The proposed method effectively and accurately identifies drilling pump faults, providing engineering value.

5.5. Domain adaptation capability of the model

To evaluate the generalization capability of the proposed method, a domain adaptation experiment across different datasets is established. The fine-tuning method facilitates the adaptation of a model, originally trained in one particular condition, to various other conditions [58]. Fig. 12 demonstrates that, post-training, the backbone of the proposed WCCN-BiLSTM method is frozen, with only the output layer remaining adjustable. Subsequently, the trained network is deployed to the training set of the target domain for a single iteration. As previously discussed, training in the target domain exclusively modifies the output layer, leaving the backbone of the proposed method unchanged. Upon completion of this training, the network is tested using the test set of the target domain to assess its adaptive capabilities.

In the domain adaptation experiments, a single dataset is chosen as the source domain for training the model, while datasets from alternative working conditions are employed as target domains to evaluate the adaptive effect of the model. Like the comparison experiments, the domain adaptation experiments unfolded on two distinct computers. During this process, meticulous attention was devoted to documenting the model training duration, as well as the time consumed for adaptive training and testing phases. Table 10 presents the specific outcomes of the experiment, where the model, initially trained using source domain data, underwent five adaptive trials. The average values of these experiments were then methodically computed and recorded. From Table 10, it is evident that adaptive training and testing significantly reduce time consumption compared to model training. Moreover, superior results were attained, with accuracy rates consistently surpassing 95%, indicating robust generalization of the proposed method.

6. Conclusions

This study proposes an intelligent fault diagnosis method, WCCN-BiLSTM, for drilling pump fault diagnosis. The method proposed offers a highly effective strategy for diagnosing faults in drilling pump. Employing the Random Forest model, identify the signal that most accurately characterizes drilling pump failures. The WCCN network, combined with BiLSTM, improves the noise reduction and feature extraction capabilities of time series data, resulting in enhanced diagnostic precision. This model is particularly effective at quickly processing noisy data for feature extraction, as well as improving features and performing sequence calculations, which lays a solid foundation for superior diagnostic performance. The domain adaptation experiments achieved impressive results, with over 95% accuracy in target domains by sequentially using each working condition's dataset as the source domain and fine-tuning the model without altering its core weights, demonstrating its strong generalization capability. However, relying solely on a single source machine may not yield sufficient knowledge for diagnostic decisions regarding the target machine. Therefore, future

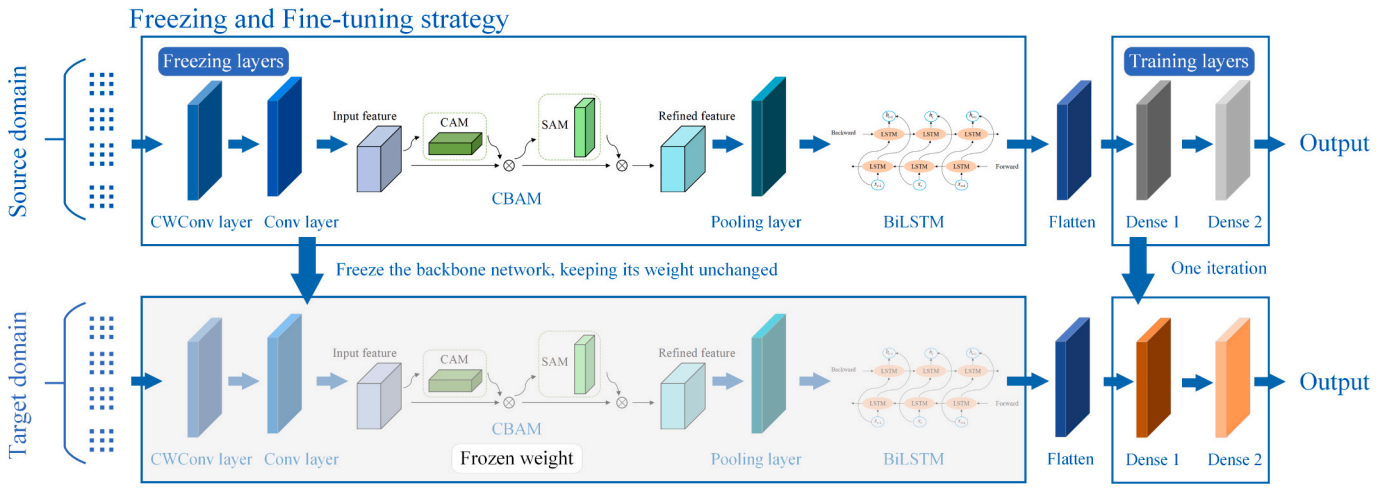


Fig. 12. Model-based transfer learning for WCCN-BiLSTM.

Table 10
Performance of the proposed method in domain adaptation experiments.

Domain Adaptation	Working Condition	Accuracy	Prediction	Recall	F1-sore	Time1	Time2
Source domain1	110 SPM	0.9913	0.9915	0.9913	0.9912	625.14 s	399.46 s
	50 SPM	0.9753	0.9777	0.9753	0.9755	14.93 s	11.37 s
Target domain1	90 SPM	0.9856	0.9860	0.9856	0.9856	14.78 s	11.53 s
	130 SPM	0.9696	0.9701	0.9696	0.9695	15.02 s	11.76 s
Source domain2	50 SPM	0.9928	0.9928	0.9927	0.9927	619.48 s	406.36 s
	90 SPM	0.9776	0.9790	0.9776	0.9777	15.12 s	11.77 s
Target domain2	110 SPM	0.9741	0.9752	0.9741	0.9741	15.03 s	11.71 s
	130 SPM	0.9594	0.9624	0.9594	0.9600	14.85 s	11.60s
Source domain3	90 SPM	0.9922	0.9923	0.9922	0.9922	637.12 s	394.81 s
	50 SPM	0.9731	0.9746	0.9731	0.9730	14.97 s	11.58 s
Target domain3	110 SPM	0.9804	0.9811	0.9804	0.9804	14.60s	11.39 s
	130 SPM	0.9578	0.9589	0.9578	0.9577	15.40s	11.71 s
Source domain4	130 SPM	0.9928	0.9929	0.9928	0.9928	624.56 s	398.56 s
	50 SPM	0.9811	0.9814	0.9811	0.9811	14.61 s	11.92 s
Target domain4	90 SPM	0.9774	0.9776	0.9774	0.9774	15.20s	11.77 s
	110 SPM	0.9786	0.9791	0.9786	0.9787	14.70s	11.56 s

research will explore the development of a multi-source diagnostic knowledge fusion module to identify faults that may not be evident within the target domain. Additionally, an important future direction involves effectively integrating the proposed method with engineering practice.

Authors Agreement

The paper has not been published as a part or as a whole, all authors agree to submit and publish it in the Applied Energy Journal.

CRediT authorship contribution statement

Junyu Guo: Methodology, Formal analysis, Conceptualization. **Yulai Yang:** Formal analysis, Investigation, Methodology, Writing – review & editing. **He Li:** Writing – original draft, Supervision, Methodology, Investigation, Formal analysis, Conceptualization. **Jiang Wang:** Writing – original draft, Visualization, Validation, Investigation, Conceptualization. **Aimin Tang:** Writing – original draft, Visualization, Validation, Investigation, Conceptualization. **Daiwei Shan:** Writing – original draft, Visualization, Validation, Investigation, Conceptualization. **Bangkui Huang:** Validation, Data curation.

Declaration of competing interest

There are no Conflicts of Interest.

Data availability

Data will be made available on request.

Acknowledgments

This research was supported by Natural Science Foundation of SiChuan, China (Grant No. 2023NSFSC0854), Joint Fund of Key Laboratory of Oil & Gas Equipment, Ministry of Education (Southwest Petroleum University) and Honghua Group Co., Ltd. (Grant No. OGEHH202005), the scholarship from China Scholarship Council (CSC) (Grant No. 202108515059).

References

[1] Calderón AJ, Pekney NJ. Optimization of enhanced oil recovery operations in unconventional reservoirs. Appl Energy 2020;258:114072.
[2] Lobo DM, Ritto TG, Castello DA. A novel stochastic process to model the variation of rock strength in bit-rock interaction for the analysis of drill-string vibration. Mech Syst Signal Pr 2020;141:106451.
[3] Liu Z, Guedes Soares C. Numerical study of rope materials of the mooring system for gravity cages. Ocean Eng 2024;298:117135.
[4] Zhang Z, Lai X, Wu M, Chen L, Lu C, Du S. Fault diagnosis based on feature clustering of time series data for loss and kick of drilling process. J Process Control 2021;102:24–33.
[5] Han Y, Liu J, Liu F, Geng Z. An intelligent moving window sparse principal component analysis-based case based reasoning for fault diagnosis: case of the drilling process. Isa T 2022;128:242–54.
[6] Ye H, Li W, Lin S, Ge Y, Lv Q. A framework for fault detection method selection of oceanographic multi-layer winch fibre rope arrangement. Measurement : Journal of the International Measurement Confederation 2024;226:114168.
[7] Barberá L, Crespo A, Viveros P, Stegmaier R. A case study of GAMM (graphical analysis for maintenance management) in the mining industry. Reliab Eng Syst Saf 2014;121:113–20.
[8] Bejger A, Piasecki T. The use of acoustic emission elastic waves for diagnosing high pressure mud pumps used on drilling rigs. Energies 2020;13:1138.
[9] Deng S, Pei J, Wang Y, Liu B. Research on drilling mud pump fault diagnosis based on fusion of acoustic emission and vibration technology. Insight - Non-Destructive Testing and Condition Monitoring 2017;59:415–23.
[10] Wang S, Chen H. A novel deep learning method for the classification of power quality disturbances using deep convolutional neural network. Appl Energy 2019;235:1126–40.
[11] Shi M, Ding C, Wang R, Song Q, Shen C, Huang W, et al. Deep hypergraph autoencoder embedding: An efficient intelligent approach for rotating machinery fault diagnosis. Knowl-Based Syst 2023;260:110172.
[12] Chen Q, Wei H, Rashid M, Cai Z. Kernel extreme learning machine based hierarchical machine learning for multi-type and concurrent fault diagnosis. Measurement 2021;184:109923.
[13] Li Z, Jiang W, Zhang S, Sun Y, Zhang S. A hydraulic pump fault diagnosis method based on the modified ensemble empirical mode decomposition and wavelet kernel extreme learning machine methods. Sensors-Basel 2021;21:2599.
[14] Liu Z, Fang L, Jiang D, Qu R. A machine-learning-based fault diagnosis method with adaptive secondary sampling for multiphase drive systems. Ieee T Power Electr 2022;37:8767–72.
[15] Lan Y, Hu J, Huang J, Niu L, Zeng X, Xiong X, et al. Fault diagnosis on slipper abrasion of axial piston pump based on extreme learning machine. Measurement 2018;124:378–85.
[16] Gao Y, Miyata S, Akashi Y. How to improve the application potential of deep learning model in HVAC fault diagnosis: based on pruning and interpretable deep learning method. Appl Energy 2023;348:121591.
[17] Wei S, Wang D, Peng Z, Feng Z. Variational nonlinear component decomposition for fault diagnosis of planetary gearboxes under variable speed conditions. Mech Syst Signal Pr 2022;162:108016.
[18] Wang C, Xin C, Xu Z. A novel deep metric learning model for imbalanced fault diagnosis and toward open-set classification. Knowl-Based Syst 2021;220:106925.
[19] Zhao B, Zhang X, Li H, Yang Z. Intelligent fault diagnosis of rolling bearings based on normalized CNN considering data imbalance and variable working conditions. Knowl-Based Syst 2020;199:105971.
[20] Kumar A, Gandhi CP, Zhou Y, Kumar R, Xiang J. Improved deep convolution neural network (CNN) for the identification of defects in the centrifugal pump using acoustic images. Appl Acoust 2020;167:107399.
[21] Tang S, Zhu Y, Yuan S. An improved convolutional neural network with an adaptable learning rate towards multi-signal fault diagnosis of hydraulic piston pump. Adv Eng Inform 2021;50:101406.
[22] Li G, Hu J, Shan D, Ao J, Huang B, Huang Z. A CNN model based on innovative expansion operation improving the fault diagnosis accuracy of drilling pump fluid end. Mech Syst Signal Pr 2023;187:109974.
[23] Tang A, Zhao W. A Fault Diagnosis Method for Drilling Pump Fluid Ends Based on Time–Frequency Transforms. Processes 2023;11:1996.
[24] Guo L, Li N, Jia F, Lei Y, Lin J. A recurrent neural network based health indicator for remaining useful life prediction of bearings. Neurocomputing 2017;240:98–109.
[25] Wang J, Dou Y, Wang Z, Jiang D. Multi-fault diagnosis method for wind power generation system based on recurrent neural network. Proceedings of the

Institution of Mechanical Engineers, Part a: Journal of Power and Energy 2019; 233:604–15.
[26] Sepp H, Jürgen S. Long short-term memory. Neural Comput 1997;8:1735–80.
[27] Bie F, Du T, Lyu F, Pang M, Guo Y. An integrated approach based on improved CEEMDAN and LSTM deep learning neural network for fault diagnosis of reciprocating pump. Ieee Access 2021;9:23301–10.
[28] Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. Arxiv 2014. arXiv preprint arXiv: 1412.3555.
[29] Miao X, Li S, Zhu Y, An Z. A novel real-time fault diagnosis method for planetary gearbox using transferable hidden layer. IEEE Sensors J 2020;20:8403–12.
[30] Han L, Deng Y, Chen H, Wei G, KaiSheng Shi J. A robust VRF fault diagnosis method based on ensemble BiLSTM with attention mechanism: considering uncertainties and generalization. Energ. Buildings 2022;269:112243.
[31] Liang X, Chen S, Zhu X, Jin X, Du Z. Domain knowledge decomposition of building energy consumption and a hybrid data-driven model for 24-h ahead predictions. Appl Energy 2023;344:121244.
[32] Azar K, Hajiakhondi-Meybodi Z, Naderkhani F. Semi-supervised clustering-based method for fault diagnosis and prognosis: a case study. Reliab Eng Syst Saf 2022; 222:108405.
[33] Li X, Li J, Qu Y, He D. Gear pitting fault diagnosis using integrated CNN and GRU network with both vibration and acoustic emission signals. Appl Sci 2019;9:768.
[34] Wang Z, Dong Y, Liu W, Ma Z. A novel fault diagnosis approach for chillers based on 1-D convolutional neural network and gated recurrent unit. Sensors-Basel 2020; 20:2458.
[35] Liao G, Gao W, Yang G, Guo M. Hydroelectric generating unit fault diagnosis using 1-D convolutional neural network and gated recurrent unit in small hydro. IEEE Sensors J 2019;19:9352–63.
[36] Zhang X, Hua X, Zhu J, Ma M. Intelligent fault diagnosis of liquid rocket engine via interpretable LSTM with multisensory data. Sensors-Basel 2023;23:5636.
[37] You K, Qiu G, Gu Y. Rolling bearing fault diagnosis using hybrid neural network with principal component analysis. Sensors-Basel 2022;22:8906.
[38] Xiang L, Wang P, Yang X, Hu A, Su H. Fault detection of wind turbine based on SCADA data analysis using CNN and LSTM with attention mechanism. Measurement 2021;175:109094.
[39] Wang J, Guo J, Wang L, Yang Y, Wang Z, Wang R. A hybrid intelligent rolling bearing fault diagnosis method combining WKN-BiLSTM and attention mechanism. Meas Sci Technol 2023;34:85106.
[40] Huang J, Mo J, Zhang J, Ma X. A Fiber vibration signal recognition method based on CNN-CBAM-LSTM. Appl Sci 2022;12:8478.
[41] Xu Z, Mei X, Wang X, Yue M, Jin J, Yang Y, et al. Fault diagnosis of wind turbine bearing using a multi-scale convolutional neural network with bidirectional long short term memory and weighted majority voting for multi-sensors. Renew Energy 2022;182:615–26.
[42] Yang Z, Zhang J, Zhao Z, Zhai Z, Chen X. Interpreting network knowledge with attention mechanism for bearing fault diagnosis. Appl Soft Comput 2020;97: 106829.
[43] You K, Qiu G, Gu Y. An efficient lightweight neural network using BiLSTM-SCN-CBAM with PCA-ICEEMDAN for diagnosing rolling bearing faults. Meas Sci Technol 2023;34:94001.
[44] Glowacz A. Ventilation diagnosis of minigrinders using thermal images. Expert Syst Appl 2024;237:121435.
[45] Peng D, Wang H, Desmet W, Gryllias K. RMA-CNN: a residual mixed-domain attention CNN for bearings fault diagnosis and its time-frequency domain interpretability. Journal of Dynamics, Monitoring and Diagnostics 2023;2:115–32.
[46] Guo J, Wang Z, Li H, Yang Y, Huang C, Yazdi M, et al. A hybrid prognosis scheme for rolling bearings based on a novel health indicator and nonlinear wiener process. Reliab Eng Syst Saf 2024;245:110014.
[47] Xiang L, Yang X, Hu A, Su H, Wang P. Condition monitoring and anomaly detection of wind turbine based on cascaded and bidirectional deep learning networks. Appl Energy 2022;305:117925.
[48] Wang C, Liu J, Zio E. A modified generative adversarial network for fault diagnosis in high-speed train components with imbalanced and heterogeneous monitoring data. Journal of Dynamics, Monitoring and Diagnostics 2022:84–92.
[49] Li T, Zhao Z, Sun C, Cheng L, Chen X, Yan R, et al. WaveletKernelNet: An interpretable deep neural network for industrial intelligent diagnosis. IEEE Trans Syst Man Cybern Syst 2022;52:2302–12.
[50] Woo S, Park J, Lee J, Kweon IS. CBAM: convolutional block attention module. Proceedings of the European Conference On Computer Vision (Eccv) 2018:3–19.
[51] Glowacz A. Thermographic fault diagnosis of electrical faults of commutator and induction motors. Eng Appl Artif Intell 2023;121:105962.
[52] Wang W, Lei Y, Yan T, Li N, Nandi A. Residual convolution long short-term memory network for machines remaining useful life prediction and uncertainty quantification. Journal of Dynamics, Monitoring and Diagnostics 2022;1:2–08.
[53] Guo J, Yang Y, Li H, Dai L, Huang B. A parallel deep neural network for intelligent fault diagnosis of drilling pumps. Eng Appl Artif Intell 2024;133:108071.
[54] Breiman L. Random forests. Mach Learn 2001;45:5–32.
[55] Tang S, Zhu Y, Yuan S. Intelligent fault identification of hydraulic pump using deep adaptive normalized CNN and synchrosqueezed wavelet transform. Reliab Eng Syst Saf 2022;224:108560.
[56] Tang S, Zhu Y, Yuan S. An adaptive deep learning model towards fault diagnosis of hydraulic piston pump using pressure signal. Eng Fail Anal 2022;138:106300.
[57] Van der Maaten L, Hinton G. Visualizing Data using t-SNE. J Mach Learn Res 2008; 9.
[58] Zhong H, Yu S, Trinh H, Lv Y, Yuan R, Wang Y. Fine-tuning transfer learning based on DCGAN integrated with self-attention and spectral normalization for bearing fault diagnosis. Measurement 2023;210:112421.