applied sciences

MDPI

Article

Incipient Fault Diagnosis in Power Cables Based on WOA-CEEMDAN and a TCN-BiLSTM Network with Multi-Head Attention

Yuhua Xing and Yaolong Yin *

School of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048, China
* Correspondence: 17839311922@163.com; Tel.: +86-17839311922

Featured Application

This study provides a practical framework for the diagnosis of incipient cable faults in power distribution systems. The proposed method is applicable to cable condition assessment in urban power grids, industrial distribution networks, and other electrical infrastructures requiring reliable offline fault analysis. It also provides a methodological basis for the future development of intelligent monitoring and early-warning systems.

Abstract

Incipient faults in power cables are difficult to diagnose because their transient signatures are weak, non-stationary, and easily masked by background noise, while labeled real-world samples are often scarce. To address these challenges, this paper proposes an offline diagnosis framework that integrates Whale Optimization Algorithm (WOA)-guided CEEMDAN with a TCN-BiLSTM-Multi-HeadAttention network. The proposed method has three main features. First, WOA is explicitly mapped to the CEEMDAN parameter optimization problem and is used to adaptively optimize the noise amplitude and ensemble number, thereby improving decomposition quality and enhancing weak fault-related components. Second, the optimized intrinsic mode functions are reconstructed into a multi-channel representation that preserves complementary fault information across different frequency bands. Third, a hybrid deep architecture combining Temporal Convolutional Networks, Bidirectional Long Short-Term Memory, and multi-HeadAttention is designed to jointly capture local transient characteristics, bidirectional temporal dependencies, and fault-sensitive feature interactions. Experimental results on both PSCAD/EMTDC simulation data and real-world measured data show that the optimized WOA-CEEMDAN achieves superior decomposition performance, with an RMSE of 0.097 and an SNR of 8.42 dB. On the real-world test dataset, the proposed framework achieves 96.00% accuracy, 97.25% precision, 96.84% recall, an F1-score of 0.970, and an AUC of 0.97, outperforming several representative baseline models. Additional ablation, noise-robustness, small-sample, confusion-matrix, and cross-cable validation results further demonstrate the effectiveness and robustness of the proposed framework for incipient cable fault diagnosis.

Keywords: cable incipient fault diagnosis; whale optimization algorithm; CEEMDAN; TCN-BiLSTM-Multihead Attention

Check for updates

Academic Editor: Roberto Zivieri

Received: 9 March 2026
Revised: 13 April 2026
Accepted: 15 April 2026
Published: 17 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

Power cables are critical components of modern power transmission and distribution systems and are widely deployed in urban distribution networks, industrial power supply

systems, offshore wind farms, and transportation infrastructure. With the rapid growth of electricity demand and the continuous expansion of power grids, the scale and complexity of cable networks have increased significantly. Consequently, the operational reliability of power cables directly affects the stability and safety of the entire power system. However, during long-term operation, cables are inevitably exposed to insulation aging, thermal stress, mechanical damage, and environmental corrosion, which may gradually degrade their insulation performance and lead to various types of faults. Among these failures, many permanent faults are preceded by incipient faults, which often manifest as intermittent discharge phenomena or weak insulation degradation before evolving into catastrophic failures [1,2]. Early detection of such faults is therefore of great importance for improving power system reliability and reducing maintenance costs.

In the past decades, numerous studies have investigated cable fault detection and location techniques. Traditional methods mainly include bridge-based techniques, impedance analysis, and traveling-wave-based fault location methods. These methods have been widely applied to locate permanent faults in power cables and distribution networks [3–5]. For example, traveling-wave-based approaches utilize high-frequency transient signals generated during fault occurrence to determine fault location with high accuracy [6]. Other approaches rely on voltage distribution characteristics or distributed parameter models to estimate fault distances in long-distance cables [7]. Although these techniques are effective for locating permanent faults, they are generally less sensitive to high-impedance or incipient faults, which often exhibit weak transient features.

To overcome this limitation, researchers have introduced advanced signal processing techniques to extract weak fault signatures from cable monitoring signals. Time–frequency analysis methods such as wavelet transform and empirical mode decomposition (EMD) have been widely applied to analyze transient signals in power systems [8,9]. However, EMD often suffers from mode mixing problems when processing complex non-stationary signals. To address this issue, improved signal decomposition algorithms such as Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) and Variational Mode Decomposition (VMD) have been proposed to enhance signal decomposition quality and improve feature extraction capability [10–12]. These methods have demonstrated strong potential in power equipment fault diagnosis by effectively separating signal components in different frequency bands.

With the rapid development of artificial intelligence technologies, machine learning and deep learning methods have become increasingly important in the field of power system fault diagnosis. Traditional machine learning algorithms such as Support Vector Machines (SVM) and Random Forests have been used to identify fault patterns and distinguish abnormal events from normal operating conditions [13,14]. In recent years, deep learning approaches including Convolutional Neural Networks (CNN), Long Short-Term Memory networks (LSTM), and attention-based models have achieved significant progress in fault diagnosis tasks due to their strong feature learning capability [15–17]. For instance, deep convolutional networks have been successfully applied to cable fault classification tasks with improved diagnostic accuracy [18]. In addition, recent studies have integrated attention mechanisms and Transformer architectures into fault diagnosis frameworks, enabling models to capture long-range temporal dependencies and enhance feature representation [19,20].

In the context of cable fault diagnosis, several recent studies have explored the integration of deep learning and advanced signal processing techniques. For example, deep learning models combined with wavelet packet transform have been used to identify faults in transmission lines with improved robustness [21]. Incremental learning-based convolutional networks have also been proposed to enhance the generalization capability

of cable fault diagnosis systems when new fault types appear [22]. Moreover, graph neural networks and self-attention mechanisms have been introduced to analyze time-series signals for cable defect detection, achieving promising results in practical monitoring scenarios [23]. Recent review studies further indicate that artificial intelligence techniques provide powerful tools for cable fault detection and localization, especially in complex power networks with large-scale monitoring data [24].

Recent studies in adjacent sensing and diagnostic fields also indicate several methodological directions that are relevant to the present problem. For example, Taguchi- and ANOVA-based optimization frameworks have been used to statistically validate parameter effects and hyperparameter settings in AI-assisted monitoring systems [25]. Such methods provide structured factor analysis and stronger statistical interpretability, particularly when the objective is to quantify the influence of a limited number of design variables. Gaussian-process-based data-driven models have further shown advantages in surrogate modelling and uncertainty-aware regression for electromagnetic characterization problems [26]. In addition, integrated sensor/electronic-circuit monitoring architectures combined with artificial intelligence have demonstrated the practical value of coupling embedded acquisition hardware with host-side intelligent diagnosis in noisy industrial environments [27]. In contrast to orthogonal-design-based statistical optimisation strategies, the present study addresses a continuous and nonlinear CEEMDAN parameter-search problem in which the decomposition quality depends on the coupled interaction between noise amplitude and ensemble number. For this reason, a population-based meta-heuristic search strategy, namely WOA, is adopted to provide adaptive global optimisation for the decomposition stage.

Despite these advances, detecting incipient cable faults remains challenging due to the weak and intermittent characteristics of early fault signals, strong environmental noise, and the limited availability of labeled fault data. Traditional signal decomposition methods often rely on manually selected parameters, which may affect decomposition performance and diagnostic accuracy. In addition, many deep learning models struggle to simultaneously capture both local transient features and long-range temporal dependencies in fault signals. Therefore, it is necessary to develop an integrated framework that combines robust signal processing techniques with advanced deep learning architectures to improve the reliability of incipient fault diagnosis.

To address these challenges, this paper proposes a diagnostic framework for incipient cable faults based on WOA-optimized CEEMDAN and a TCN-BiLSTM-Multi-HeadAttention network. First, the Whale Optimization Algorithm is employed to adaptively optimize the key parameters of CEEMDAN, thereby improving decomposition quality and enhancing the extraction of weak fault features. Subsequently, the decomposed intrinsic mode functions are reconstructed as multi-channel inputs and fed into a hybrid deep learning architecture that integrates Temporal Convolutional Networks, Bidirectional Long Short-Term Memory, and multi-HeadAttention for fault identification. Experimental results on both simulation data and real cable-monitoring data demonstrate that the proposed method achieves superior diagnostic accuracy, robustness, and practical applicability compared with several representative approaches. The overall workflow of the proposed research framework for incipient cable fault diagnosis is shown in Figure 1.

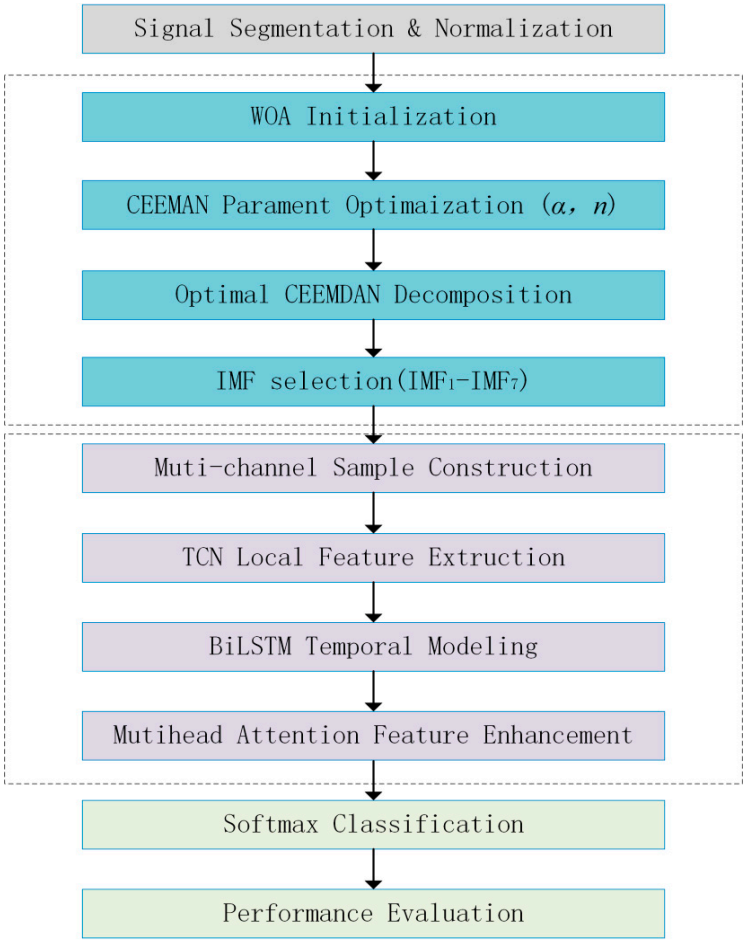


Figure 1. Overall workflow of the proposed incipient fault diagnosis framework for power cables (Blue indicates the WOA-CEEMDAN optimization module, purple represents the hybrid model, gray denotes the input, and yellow corresponds to the output).

The main contributions of this study are summarized as follows:

(1) An adaptive WOA-guided CEEMDAN framework is developed to optimize the noise amplitude and ensemble number automatically, which improves decomposition quality and enhances the extraction of weak incipient-fault components from noisy cable signals.

(2) A hybrid TCN-BiLSTM-Multihead-Attention network is constructed to jointly learn local transients, bidirectional temporal dependencies, and globally important feature interactions, thereby improving diagnostic discriminability for subtle fault patterns.

(3) An integrated signal-processing-to-diagnosis pipeline is established by explicitly mapping the optimized CEEMDAN outputs to multi-channel deep features, while the computational burden is controlled through offline parameter optimization, limited IMF retention, and a compact sequential architecture.

(4) Extensive experiments on both simulated and real measured datasets, together with ablation, small-sample, noise-robustness, ROC/AUC, and confusion-matrix analyses, verify the superiority and engineering applicability of the proposed method for incipient cable fault diagnosis.

The offline phase includes signal acquisition from both simulation and real cable-monitoring systems, signal segmentation and normalization, WOA-based optimization of CEEMDAN parameters, adaptive signal decomposition, IMF selection and multi-channel feature construction, model training, and performance validation. The subsequent stage

illustrates the intended diagnosis procedure for future engineering implementation, including new signal acquisition, fixed-parameter CEEMDAN decomposition, IMF-based feature reconstruction, TCN-based local feature extraction, BiLSTM-based temporal modeling, multi-head-attention-based feature reweighting, Softmax-based fault classification, and alarm output. It should be noted that, in the present study, this procedure was validated as an offline diagnosis framework based on field-acquired data rather than as a fully online embedded deployment.

2. WOA-Optimized CEEMDAN

2.1. Whale Optimization Algorithm

WOA simulates the hunting behavior of humpback whales through three phases: encircling prey, bubble-net attacking, and prey search [28–30]. Each whale represents a candidate solution in the search space. In the encircling phase, position updates follow (1):

$$\overline{D} = |C \cdot X_*(t) - X_i(t)|,\ X_i(t+1) = X_*(t) - A \cdot D \tag{1}$$

where $X_i(t)$ is the position of the i-th search agent at iteration t, $X_*(t)$ is the global best position, and A and C are coefficient vectors.

The bubble-net attacking phase combines shrinking encircling and spiral updating (2):

$$X(t+1) = \overline{D} \times e^{lb} \times \cos(2\pi t) + X_*(t) \tag{2}$$

where b is a constant defining the logarithmic spiral shape, and l is a random number in $[-1, 1]$. During prey search (exploration phase), when $|A| > 1$, agents are pushed apart:

$$X(t+1) = \overline{X_{rand}} - A \cdot \left|\overline{C} \cdot \overline{X_{rand}} - X(t)\right| \tag{3}$$

where $\overline{X_{rand}}$ is a randomly selected search agent position.

2.2. Complete Ensemble Empirical Mode Decomposition with Adaptive Noise

Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) suppresses mode mixing and random noise, enabling adaptive signal decomposition [31]. This method decomposes the signal into several Intrinsic Mode Functions (IMFs), each containing feature information from different frequency bands, effectively extracting local mutations and key components from the signal. The core steps are as follows:

- Add white noise to the original signal $x(t)$ across multiple realizations:

$$x_i(t) = x(t) + \varepsilon_0 \beta_i(t),\ i = 1, 2 \ldots N \tag{4}$$

where N denotes the number of noise realizations, ε_0 is the noise amplitude coefficient, and i indexes the i-th noisy signal.

- Decompose each noisy signal using Empirical Mode Decomposition (EMD) to obtain the first IMF, then compute the ensemble average:

$$IMF = \frac{1}{N}\sum_{i=1}^{N} EMD(x(t) + \varepsilon_0 \beta_i(t)) \tag{5}$$

where $EMD(t)$ represents the EMD operation.

- After extracting the k-th IMF, the corresponding residual is given by:

$$r_k(t) = x(t) - \sum_{j=1}^{k} IMF_j(t) \tag{6}$$

where $IMF_j(t)$ denotes the ensemble average of the j-th IMF.

- For the (k + 1)-th IMF, add adaptive noise to the current residual and repeat the decomposition process until the stopping criterion (residual extrema ≤ 1) is met:

$$r_k^i(t) = r_k(t) + \varepsilon_k \beta_i(t) \tag{7}$$

$$IMF_{k+1} = \frac{1}{N}\sum_{i=1}^{N} EMD(r_k^i(t)) \tag{8}$$

where ε_k is the noise amplitude coefficient at the k-th stage, and $\beta_i(t)$ is the added noise realization.

- The original signal can be perfectly reconstructed by summing all IMF components and the final residual:

$$x(t) = \sum_{i=1}^{m} IMF_i(t) + r(t) \tag{9}$$

where m is the total number of IMFs obtained from the decomposition.

2.3. WOA-CEEMDAN Optimization Process

The performance of CEEMDAN is highly dependent on the selection of the noise amplitude and ensemble number. In this study, WOA is employed to optimize these two critical parameters. The optimization procedure is illustrated in Figure 2 and can be summarized as follows:

(1) Initialize WOA parameters: Population size N ϵ [20, 50], maximum iterations [30, 100], spiral constant b ϵ [0.5, 1.5].
(2) Define optimization vector: p = [α, n], where α ϵ [0.05, 0.3] is noise amplitude and n ϵ [50, 150] is ensemble number.
(3) Perform CEEMDAN decomposition on cable incipient fault signal x(t) using candidate parameters.
(4) Calculate fitness function based on spectral entropy of each IMF:

$$H_i = -\sum_{k} \widetilde{P}_i(k) \log_2 \widetilde{P}_i(k) \tag{10}$$

where H_i is the spectral entropy of the i-th IMF, $\widetilde{P}_i(k)$ is the normalized power spectrum.

(5) Construct fitness function F(P) to minimize noise IMF influence while preserving fault features:

$$F(P) = \sum_{I=1}^{K} w_i H_i \tag{11}$$

where w_i is the weight for the i-th IMF.

(6) Update whale positions based on current optimal solution and repeat steps (3)–(5).
(7) Output global optimal parameters upon reaching maximum iterations.

It should be noted that alternative optimisation strategies, including statistically structured approaches such as Taguchi- and ANOVA-based frameworks, can provide stronger interpretability for factor-effect analysis and are valuable when the optimisation problem involves a limited number of discretised design levels [30]. However, in the present study, the CEEMDAN parameter optimisation problem is continuous, nonlinear, and strongly coupled to the decomposition-quality objective. In particular, the joint influence of the noise amplitude and ensemble number on mode separation quality is difficult to characterize using a fixed orthogonal design alone. Therefore, WOA is adopted in this work because it offers a flexible population-based global search mechanism for jointly optimising these parameters in a data-adaptive manner. Nevertheless, statistically structured optimisation

and validation strategies remain meaningful complementary tools and will be incorporated in future work to further enhance the interpretability of the optimisation process.

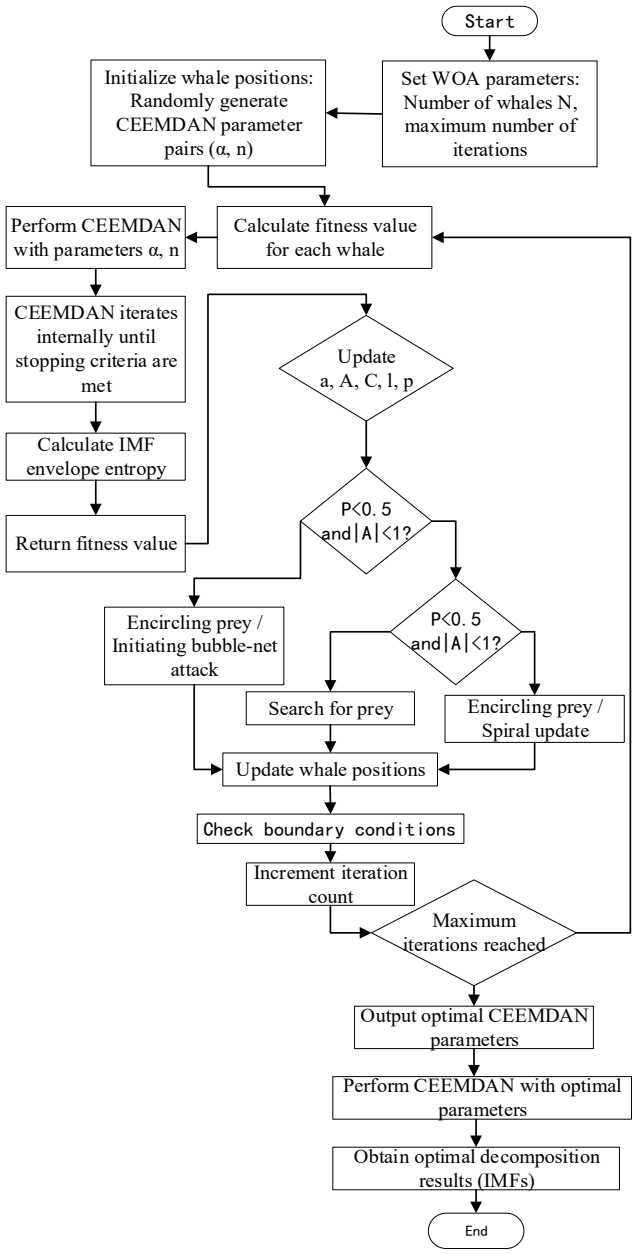


Figure 2. Flowchart of the WOA-based optimization process for CEEMDAN parameter selection.

Table 1 summarizes the main parameter settings adopted in the proposed WOA-CEEMDAN framework. Following the optimisation strategy described in Section 2.3, the WOA population size, maximum iteration number, and spiral constant were fixed, whereas the CEEMDAN amplitude and ensemble number were jointly searched within predefined continuous ranges. After optimisation, the best parameter combination was obtained as $\alpha = 0.13$ and $n = 126$, with the maximum decomposition level set to 7 and IMF1–IMF6 retained for subsequent feature construction. Unlike Taguchi–ANOVA-based statistical optimisation frameworks, which typically discretise factors into a limited number of predefined levels, the present study formulates CEEMDAN parameter tuning as a continuous and coupled search problem. Therefore, WOA was adopted to provide adaptive global optimisation over the continuous parameter space while still allowing the parameter settings to be clearly summarised in a structured tabular form.

Table 1. Parameter settings of the proposed WOA-CEEMDAN optimisation.

Symbol	Parameter	Setting in This Study
N_{pop}	Population size	30
$T_{\max}$	Maximum iterations	100
b	Spiral constant	1.0
α	Noise amplitude	[0.05, 0.30]
n	Ensemble number	[50, 150]
L	Maximum decomposition level	7

2.4. Problem Formulation and Mapping of WOA to the Diagnostic Task

In this study, WOA is not used directly for fault classification; instead, it is employed as a parameter-search strategy for CEEMDAN so that the subsequent diagnosis model receives more informative and less noisy inputs. Specifically, each whale represents one candidate CEEMDAN parameter vector, i.e.,

$$P_i = [\alpha_i, n_i] \tag{12}$$

where α_i denotes the added noise amplitude and n_i denotes the ensemble number. For a given cable current signal, the candidate vector P_i s first used to perform CEEMDAN decomposition, yielding a set of IMFs $\{c_1(t), c_2(t) \ldots\ldots c_m(t)\}$. The quality of this decomposition is then evaluated using the fitness function defined in (11), which favors IMF sets with lower spectral entropy and clearer fault-related structure. Hence, the search space of WOA corresponds to the CEEMDAN parameter space, while the objective of WOA is to find the parameter combination that maximizes fault-feature separability before classification.

After optimization, the CEEMDAN parameters are fixed at their optimal values and the resulting IMF components are reorganized as multi-channel inputs for the diagnosis network. In this way, the WOA layer addresses parameter adaptivity in signal decomposition, whereas the TCN-BiLSTM-Multihead-Attention network performs feature learning and fault identification. This explicit division of roles establishes a clear mapping between WOA optimization and the incipient fault diagnosis task.

3. TCN-BiLSTM-Multihead-Attention Model

To effectively capture both local and global temporal dependencies in cable incipient fault signals, this paper proposes a hybrid deep learning model that integrates Temporal Convolutional Network (TCN) [32], Bidirectional Long Short-Term Memory (BiLSTM), and Multihead Attention mechanism. This architecture leverages the strengths of each component to enhance feature extraction and sequence modeling capabilities.

3.1. Temporal Convolutional Network (TCN)

TCN is a variant of convolutional networks designed for sequence modeling tasks. It employs causal convolutions to ensure that no future information leaks into the past, making it suitable for time-series analysis. The architecture also incorporates dilated convolutions to exponentially increase the receptive field without significantly increasing the number of parameters. This allows TCN to capture long-range dependencies with high computational efficiency.

Given an input sequence $X = \{x_1, x_2, \ldots x_T\}$, the output of a dilated causal convolution at time t is defined as:

$$F(t) = \sum_{i=0}^{k-1} f(i) \cdot x_{t-di} \tag{13}$$

where k is the kernel size, d is the dilation factor, and f is the convolution filter. By stacking multiple layers with increasing dilation rates, TCN captures multi-scale temporal features critical for identifying incipient fault patterns.

3.2. Bidirectional Long Short-Term Memory

While TCN excels at capturing local patterns, BiLSTM is introduced to model bidirectional temporal dependencies [33]. LSTM units mitigate the vanishing gradient problem through gating mechanisms, enabling the network to retain long-term memory. The bidirectional extension processes the sequence in both forward and backward directions, allowing the model to access past and future context simultaneously. The structural diagram is shown in Figure 3.

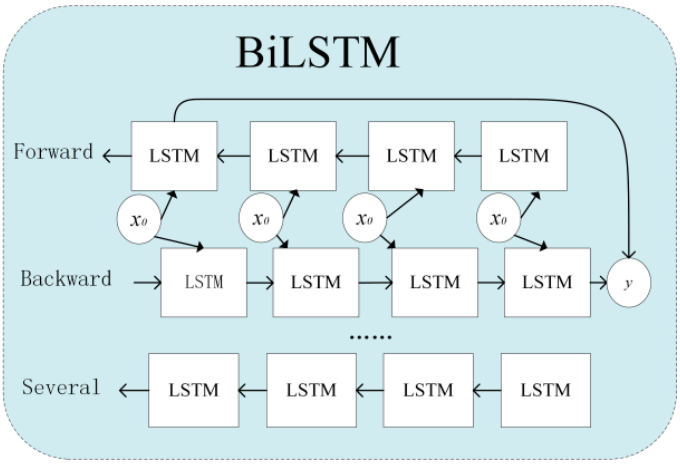


Figure 3. Schematic structure of the BiLSTM module.

For each time step t, the forward LSTM generates a hidden state $\overrightarrow{h_t}$, and the backward LSTM generates $\overleftarrow{h_t}$. These are concatenated to form the final hidden representation:

$$h_t = \left[\overrightarrow{h_t}; \overleftarrow{h_t}.\right] \tag{14}$$

This structure enhances the model's sensitivity to subtle variations in fault signals that may precede or follow a fault event.

3.3. Multihead Attention Mechanism

To further improve the model's ability to focus on informative time steps, a multi-HeadAttention mechanism is applied to the output of the BiLSTM layer. The attention mechanism assigns different weights to different time steps, enabling the model to emphasize fault-related features while suppressing irrelevant fluctuations and residual noise, as shown in Equation (15). Here, Q, K and V denote the query, key, and value matrices derived from the BiLSTM outputs, respectively, and d_k denotes the dimension of the key vectors [34]. Multi-HeadAttention performs multiple attention operations in parallel, each of which learns a different representation subspace, and then concatenates the outputs, as shown in Equations (16) and (17). By operating in multiple representation subspaces, the Multi-Head Attention mechanism enhances the model's ability to capture complex fault patterns at different temporal scales.

$$Attention(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})\text{V} \tag{15}$$

$$MultiHead(Q, K, V) = Concat(head_1, \ldots\ldots head_h)W^O \quad (16)$$

$$head_i = \text{Softmax}(\frac{QW_i^Q(KW_i^K)^T}{\sqrt{d_k}})(VW_i^V) \quad (17)$$

where W_i^Q, W_i^K, W_i^V and W^O are trainable projection matrices. By performing attention operations in multiple representation subspaces, the Multihead Attention mechanism enhances the model's ability to capture complex fault patterns at different temporal scales. For clarity, the main symbols used in Equations (15)–(17) are summarized in Table 2.

Table 2. Nomenclature for Equations.

Symbol	Meaning
Q	Query matrix derived from the BiLSTM output features
K	Key matrix derived from the BiLSTM output features
V	Value matrix derived from the BiLSTM output features
d_k	Dimension of the key vectors
Softmax	Normalization function used to obtain attention weights
QK^T	Similarity matrix between queries and keys
$head_i$	Output of the i-th attention head
W^O	Output projection matrix after concatenating all attention heads
$Concat$	Concatenation of the outputs from all attention heads

This mechanism enhances the model's capacity to capture complex fault patterns across different temporal scales.

3.4. TCN-BiLSTM-Multihead-Attention

The TCN-BiLSTM-Multihead-Attention hybrid neural network model proposed in this paper is an end-to-end deep learning architecture designed specifically for cable incipient fault signal characteristics. Its overall structure adopts a hierarchical progressive design as shown in Figure 4, primarily consisting of five core modules: the input layer receives multi-dimensional time-series data reconstructed by WOA-CEEMDAN (feature matrix composed of IMF1-IMF6); the TCN feature extraction module serves as the front end, capturing local temporal patterns and multi-scale features of fault signals through multi-layer dilated causal convolutions; the BiLSTM temporal modeling module, composed of forward and backward LSTMs, performs bidirectional modeling on the feature sequences output by TCN, simultaneously capturing past and future contextual dependencies; the Multihead-Attention feature enhancement module maps the BiLSTM outputs into query, key, and value matrices, highlighting critical time steps of fault features from different representation subspaces through parallel computation of multiple attention heads; the output layer, through global average pooling, fully connected layers, and a Softmax classifier, ultimately outputs the cable incipient fault identification results. This hybrid architecture fully leverages the local feature extraction capability of TCN, the temporal modeling capability of BiLSTM, and the feature enhancement capability of the attention mechanism, achieving efficient extraction and accurate recognition of weak cable incipient fault characteristics.

As illustrated in Figure 4, dropout layers with a rate of 0.3 were inserted after the Multihead Attention modules to reduce overfitting and improve the generalization capability of the proposed network. In the overall framework, WOA-CEEMDAN serves as the signal preprocessing module and is responsible for adaptively decomposing noisy cable current signals into informative intrinsic mode functions, thereby enhancing weak fault-related components. The TCN module acts as the front-end feature extractor, capturing local transient patterns and multi-scale temporal characteristics through dilated causal convolutions. The BiLSTM module further models bidirectional temporal dependencies, enabling the net-

work to exploit both past and future contextual information in the reconstructed sequences. The Multihead Attention module adaptively reweights the learned temporal features and emphasizes fault-sensitive representations from different subspaces, which helps suppress irrelevant fluctuations and residual noise. The dropout modules are introduced to improve model robustness and alleviate overfitting during training. Finally, the fully connected layer and Softmax classifier map the high-level features into fault categories and generate the final diagnosis results.

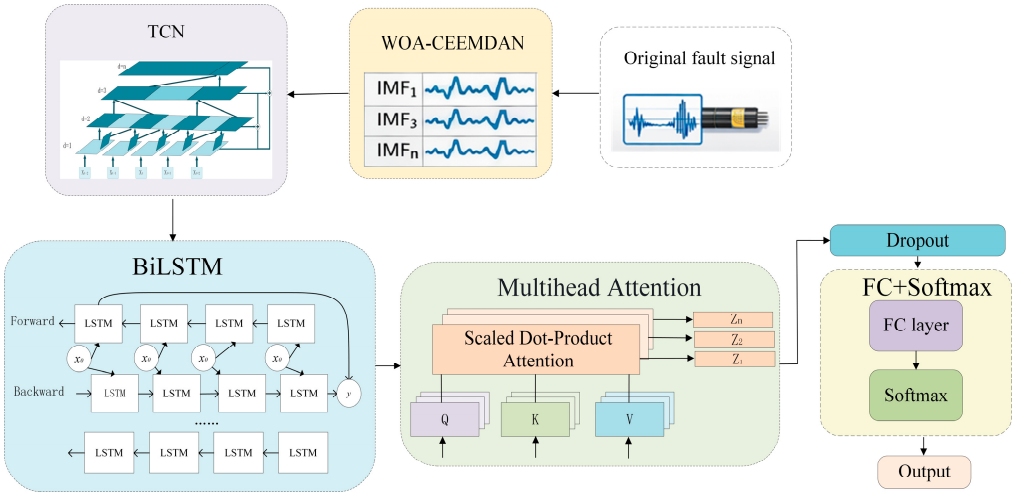


Figure 4. Overall architecture of the proposed TCN-BiLSTM-Multi-HeadAttention fault diagnosis network.

3.5. Computational Complexity and Practical Deployment Considerations

Although the proposed model combines TCN, BiLSTM, and Multi-Head Attention, its computational burden is controlled from both the preprocessing and network-design perspectives. First, WOA is used offline to determine the CEEMDAN parameters only once for a given dataset and therefore does not increase the online inference cost. Second, only the most informative IMF components are retained after decomposition, which reduces redundant input channels and suppresses noise propagation into the classifier. Third, the TCN front end extracts local features using shared convolution kernels and dilated convolutions, allowing the receptive field to be enlarged without substantially increasing the model complexity. The subsequent BiLSTM and Multi-Head Attention modules operate on compressed sequential features rather than on raw waveforms, which further reduces the sequence-modeling burden. In the present study, the framework was validated as an offline diagnosis method based on field-acquired data rather than as a fully deployed embedded system. Therefore, the current complexity discussion is intended to demonstrate methodological feasibility, whereas dedicated runtime benchmarking and real-time embedded implementation will be investigated in future work.

Formally, for an input sequence of length T, the dominant costs of the three modules can be expressed as $O(TkC^2)$, $O(TH^2)$, $O(hT^2d_h)$, respectively, where k is the kernel size, C is the channel dimension, H is the hidden size, hhh is the number of attention heads, and $T' < T$ denotes the reduced sequence length after TCN feature extraction. Therefore, the proposed hybrid architecture improves representation capability without causing an unacceptable increase in online diagnostic complexity. In addition, it should be emphasized that the current implementation corresponds to an offline diagnosis framework based on practical field-acquired data, rather than a fully online embedded diagnosis system. Accordingly, the present computational-complexity discussion is intended to demonstrate the method-

ological feasibility of the proposed architecture, while dedicated runtime benchmarking and fully embedded real-time deployment are beyond the scope of the present study.

4. Experimental Results and Analysis

4.1. Detailed Configuration Parameters

For signal preprocessing, both simulated and measured current signals were segmented into samples of 2048 points and normalized using z-score standardization. The CEEMDAN search ranges were set to $\alpha \in [0.05, 0.30]$ and $n \in [50, 150]$. The WOA population size was set to 30, the maximum iteration number was 100, and the spiral constant b was fixed to 1.0. According to the optimization results, the optimal CEEMDAN parameters were $\alpha = 0.13$ and $n = 126$, the maximum decomposition level was set to 7. IMF1–IMF6 were retained to construct the multi-channel input features for the diagnosis model.

The TCN module comprised three residual blocks with a kernel size of 3 and dilation factors of 1, 2, and 4, respectively. The numbers of convolution channels were set to 32, 64, and 64. The BiLSTM hidden size was 64 in each direction. The Multihead Attention module used 4 heads with a model dimension of 128. A dropout rate of 0.3 was applied after the BiLSTM and attention layers to alleviate overfitting. The classifier consisted of a fully connected layer with 64 neurons and a Softmax output layer.

For network training, the Adam optimizer was employed with an initial learning rate of 1×10^{-3} and a batch size of 64. The maximum training epoch was 100, and early stopping with a patience of 10 epochs was adopted according to the validation loss. The loss function was categorical cross-entropy. All experiments were conducted under the same training/validation/testing split ratio of 70%/15%/15% to ensure a fair comparison.

4.2. Dataset and Experimental Setup

To validate the effectiveness of the proposed TCN-BiLSTM-Multihead-Attention model under controlled conditions, a simulation model of a 10 kV cable distribution network was constructed using PSCAD/EMTDC as shown in Figures 5 and 6. The cable model was based on the YJV42-8.7/10 kV parameters, with a total length of 5 km. Incipient faults were simulated by introducing intermittent single-phase-to-ground arcs with fault resistance ranging from 100 Ω to 500 Ω (step 100 Ω), and fault inception angles varying from 0° to 90° (step 15°). The sampling frequency was set to 10 MHz to match the real acquisition system. A total of 12,000 simulated samples were generated, each containing 2048 time points (approximately 0.2 ms duration) covering the pre-fault, fault inception, and post-fault transient periods.

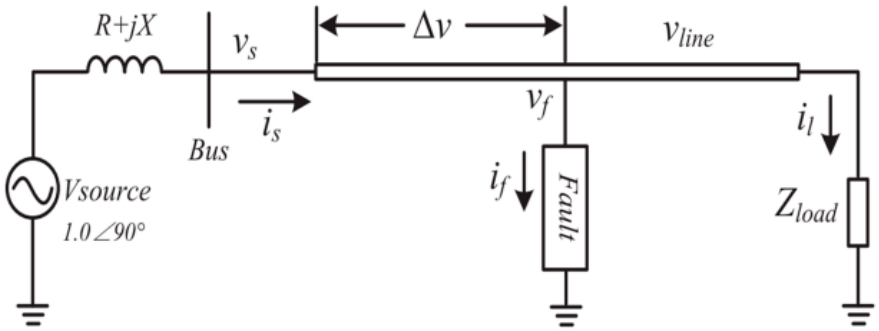


Figure 5. Circuit simulation model diagram.

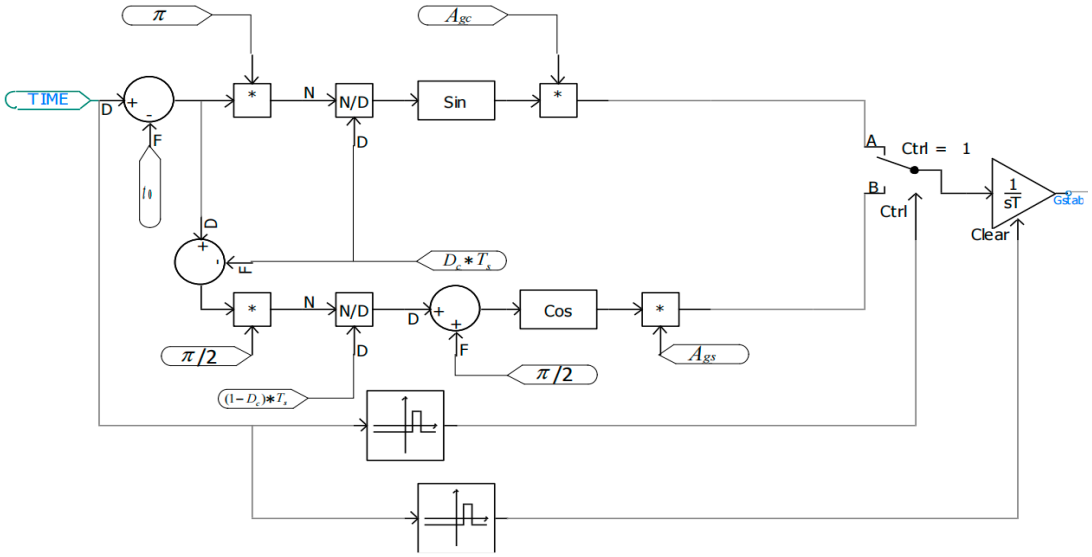


Figure 6. Arc−generation circuit used in the simulation model. Arrows indicate the signal flow direction. The symbol "*" denotes multiplication operations. Blocks represent functional modules, including logic operations, trigonometric functions (Sin/Cos), and control units. Letters such as N, D, F, and T denote intermediate variables or control parameters within the arc model, while "Ctrl" indicates control signals for switching operations.

To illustrate the behavior of incipient faults under controlled simulation conditions, Figure 7 presents the current waveform during a typical discharge event. The waveform exhibits high-frequency transient components at the fault inception moment, followed by damped oscillations, which are characteristic of intermittent arcing faults in cable systems. Correspondingly, Figure 8 shows the voltage waveform during the same discharge event. The voltage signal exhibits a sudden drop at the fault instant, accompanied by high-frequency distortion and subsequent recovery, which are important indicators of incipient fault behavior.

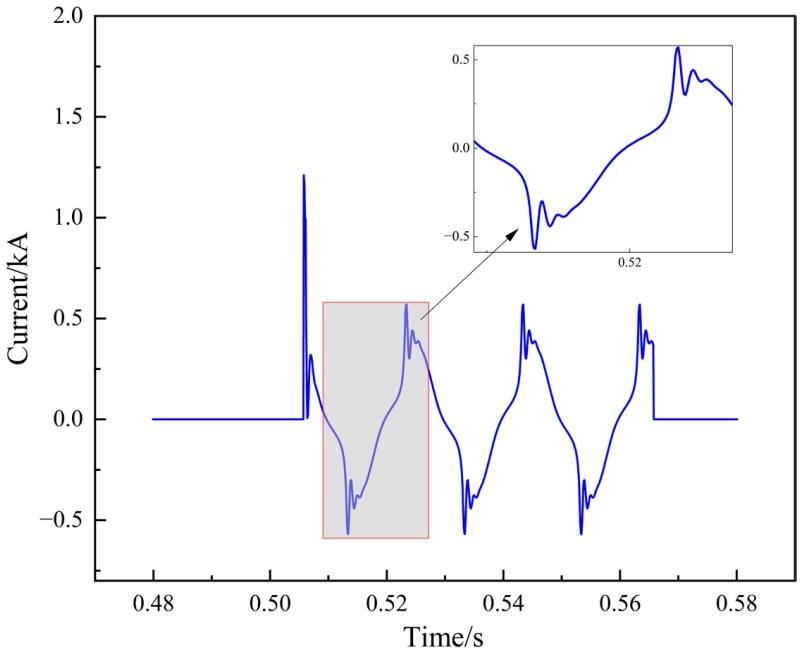


Figure 7. Current waveform during a typical discharge event.

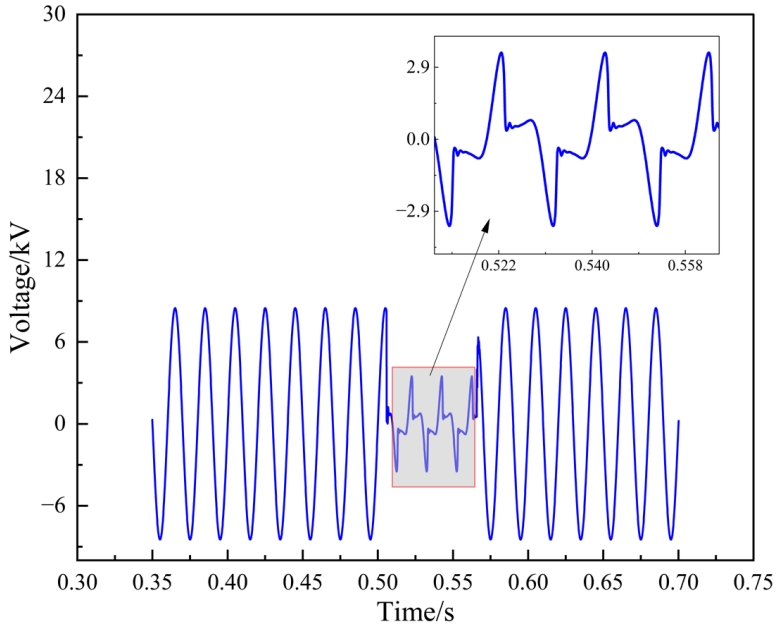


Figure 8. Voltage waveform during a typical discharge event.

To further validate the practical effectiveness of the proposed algorithm, real-world cable data were also collected and employed in this study. Experimental data were collected from an online cable fault acquisition system comprising high-frequency current sensors, a 14-bit ADC with 10 MHz sampling rate, and an MCU control unit. The test object was YJV42-8.7/10 kV cable. As shown in Figures 9 and 10, a total of 14,000 samples were collected over 2 h, randomly partitioned into training (70%), validation (15%), and testing (15%) sets.

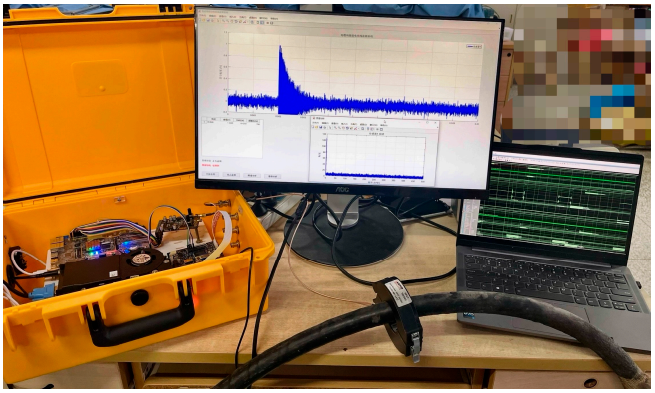


Figure 9. Experimental setup for cable-fault data acquisition.

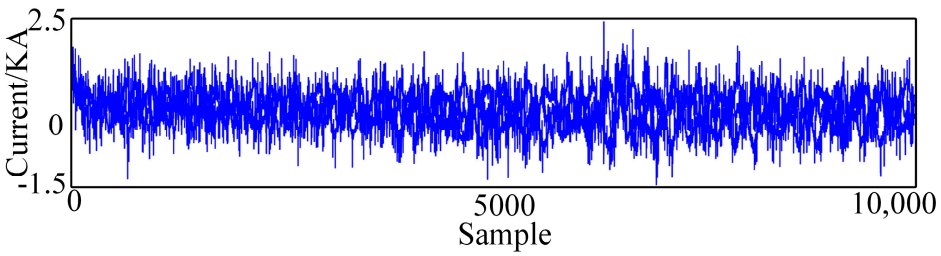


Figure 10. Measured current signals contaminated by noise and disturbances.

It should be noted that the present study focuses on offline fault diagnosis rather than fully online embedded inference. In the experimental setup, the MCU control unit was

mainly used for signal acquisition, control, and data transmission, while the subsequent CEEMDAN-based preprocessing and TCN-BiLSTM-Multihead-Attention diagnosis were performed offline on the host-side computing platform. Therefore, the purpose of the real measured dataset in this work is to validate the diagnostic effectiveness, robustness, and engineering applicability of the proposed method under practical acquisition conditions, rather than to demonstrate a fully deployed real-time edge implementation. Quantitative benchmarking of end-to-end execution time, memory usage, and embedded real-time deployment efficiency will be investigated in future work.

4.3. WOA-CEEMDAN Optimization

To further investigate the parameter sensitivity and optimization effectiveness of the proposed WOA-CEEMDAN framework, additional experiments were conducted focusing on parameter selection, decomposition performance comparison, and robustness analysis.

4.3.1. Parameter Sensitivity Analysis of WOA-CEEMDAN

The core of the WOA lies in simulating the intelligent foraging behavior of whales. To find the optimal parameter combination for CEEMDAN, this paper designs experiments to demonstrate the optimization process of the proposed scheme.

Figure 11 shows the convergence behavior of the WOA optimization process. The fitness value decreases rapidly during the early iterations and gradually stabilizes after approximately 50 iterations, reaching F(P) = 0.124. This result indicates that the proposed optimization procedure converges efficiently to a stable solution within the preset iteration range.

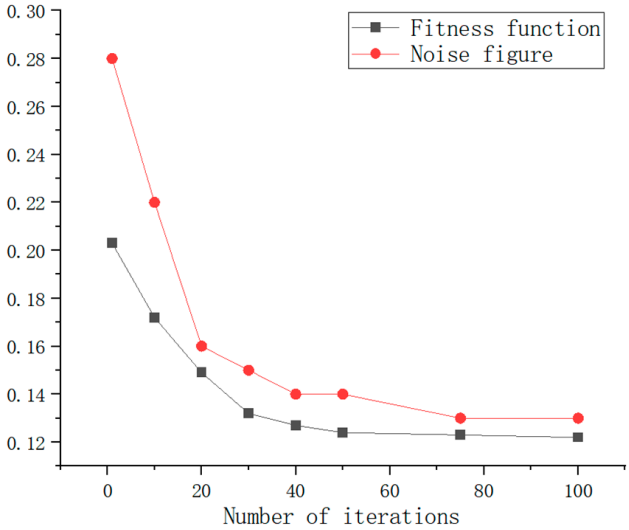


Figure 11. Convergence curve of the WOA optimization process.

The performance of CEEMDAN is highly dependent on two critical parameters: noise amplitude (α) and ensemble number (n). In order to verify the effectiveness of the WOA optimization process, multiple combinations of these parameters were tested before optimization.

The noise amplitude α was varied from 0.05 to 0.30, while the ensemble number n ranged from 50 to 150. RMSE, SNR, and information entropy were used as evaluation indicators. Table 3 presents the influence of different parameter combinations on decomposition quality.

Table 3. Sensitivity analysis of different CEEMDAN parameter combinations.

α	n	RMSE	SNR (dB)	Information Entropy (Bit)
0.05	50	0.134	5.11	16.82
0.05	100	0.128	5.43	16.10
0.05	150	0.126	5.62	15.75
0.10	50	0.120	6.12	14.80
0.10	100	0.112	6.78	13.96
0.10	150	0.109	7.01	13.52
0.13	126	0.097	8.42	12.80
0.20	100	0.105	7.32	13.15
0.25	120	0.110	6.95	13.70
0.30	150	0.118	6.41	14.22

To avoid ambiguity, the search range of α in both the WOA optimization and the pre-optimization sensitivity analysis was fixed to [0.05, 0.30]. Therefore, the α = 0.05 rows in Table 3 are not additional baseline tests outside the optimization boundary, but part of the same candidate interval used to verify parameter sensitivity before selecting the optimal solution.

The results demonstrate that excessive noise amplitude leads to noise amplification in the decomposition results, while too small noise amplitude fails to eliminate mode mixing effectively. The optimal parameter combination (α = 0.13, n = 126) identified by WOA achieves the lowest RMSE and highest SNR, indicating superior signal reconstruction and noise suppression capability.

4.3.2. Comparison of Different Objective Functions

To further examine whether the spectral-entropy-based objective function used in Section 2.3 is appropriate for CEEMDAN parameter optimisation, additional comparative experiments were conducted using several alternative objective criteria. Under the same WOA settings (population size = 30, maximum iterations = 100, and identical search ranges for α and n), four objective functions were tested: (1) spectral entropy (SE), (2) reconstruction error (RE), (3) orthogonality index (OI), and (4) a composite objective combining spectral entropy and reconstruction error. For each objective function, WOA was used to determine the optimal CEEMDAN parameters, and the resulting decomposition performance was evaluated using RMSE, SNR, and information entropy.

As shown in Table 4, different objective functions lead to different optimal CEEMDAN parameter combinations and noticeably different decomposition results. Among the tested criteria, the spectral-entropy-based objective function achieves the best overall performance, with the lowest RMSE (0.097), the highest SNR (8.42 dB), and the lowest information entropy (12.80). This indicates that spectral entropy is more effective in suppressing noisy IMF components while preserving weak fault-related transient features. By contrast, the reconstruction-error-based objective and the orthogonality-index-based objective result in higher RMSE values, lower SNR values, and larger information entropy, suggesting that optimizing only signal fidelity or IMF independence is insufficient to achieve the best fault-oriented decomposition quality.

The composite objective (SE + RE) provides intermediate performance, yielding better results than RE and OI alone, but still remaining inferior to the spectral-entropy-based objective in all three evaluation indices. This suggests that introducing reconstruction-error information can improve the balance of the optimisation to some extent, but it does not outperform spectral entropy alone for the present problem. Overall, Table 4 demonstrates that the spectral-entropy-based objective function offers the most suitable trade-off between noise

suppression, decomposition compactness, and weak-feature preservation, and is therefore adopted as the default fitness function in the proposed WOA-CEEMDAN framework.

Table 4. Comparison of different objective functions for WOA-based CEEMDAN optimisation.

Objective Function	[α, n]	RMSE	SNR (dB)	IE
Spectral entropy (SE)	[0.13, 126]	0.097	8.42	12.80
Reconstruction error (RE)	[0.21, 108]	0.112	7.15	14.36
Orthogonality index (OI)	[0.18, 95]	0.118	6.83	15.27
Composite objective (SE + RE)	[0.15, 118]	0.105	7.64	13.58

4.3.3. Comparison with Other Signal Decomposition Methods

To evaluate the effectiveness of the proposed method in early cable fault diagnosis, this study compares it with several widely used signal decomposition methods, including EMD, CEEMDAN, PSO-CEEMDAN [35], GWO-CEEMDAN, SSA-CEEMDAN, and WOA-CEEMDAN, focusing on feature extraction capability, noise suppression performance, and decomposition accuracy. To provide a comprehensive evaluation, two groups of comparative experiments are designed based on simulation data and real cable monitoring data, respectively. Three quantitative evaluation metrics, namely Root Mean Square Error (RMSE), Signal-to-Noise Ratio (SNR), and total information entropy, are employed to assess the performance of different decomposition methods. The detailed results are presented in Table 5.

Table 5. Performance comparison of different signal decomposition methods on simulation and real datasets.

Group	Method	RMSE	SNR (dB)	IE
Sim data [1]	EMD	0.142	5.12	18.20
	EEMD	0.136	5.46	17.42
	CEEMDAN	0.128	5.87	15.50
	VMD	0.121	6.73	17.92
	PSO-CEEMDAN	0.113	7.11	13.20
	GWO-CEEMDAN	0.099	8.13	12.98
	SSA-CEEMDAN	0.106	7.86	13.07
	WOA-CEEMDAN	0.093	8.42	12.80
Rel data [2]	EMD	0.158	4.83	19.31
	EEMD	0.162	5.11	18.20
	CEEMDAN	0.137	5.54	16.73
	VMD	0.159	6.12	17.84
	PSO-CEEMDAN	0.118	6.85	14.26
	GWO-CEEMDAN	0.117	7.24	13.66
	SSA-CEEMDAN	0.121	6.67	14.39
	WOA-CEEMDAN	0.109	7.58	13.61

[1] Simulation Data in Section 4.1. [2] Real-World Measured Data in Section 4.1.

The results in Table 5 show that the proposed WOA-CEEMDAN method achieves the best overall decomposition performance on both the simulation and real-world datasets. On the simulation dataset, WOA-CEEMDAN obtains the lowest RMSE (0.093), the highest SNR (8.42 dB), and the lowest information entropy (12.80), indicating superior reconstruction accuracy, stronger noise suppression, and better concentration of fault-related information. Among the other meta-heuristic optimization methods, GWO-CEEMDAN ranks second, followed by SSA-CEEMDAN and PSO-CEEMDAN, while the conventional decomposition methods, including EEMD, CEEMDAN, and VMD, show relatively inferior performance.

A similar trend can be observed on the real-world dataset. WOA-CEEMDAN again achieves the best performance, with the lowest RMSE (0.109), the highest SNR (7.58 dB), and the lowest IE (13.61). Compared with PSO-CEEMDAN, the proposed WOA-CEEMDAN further reduces the reconstruction error and improves the signal-to-noise ratio, while also producing more concentrated and informative intrinsic mode components. GWO-CEEMDAN provides competitive results on the real dataset and ranks second overall, whereas SSA-CEEMDAN also improves upon conventional CEEMDAN but remains inferior to both GWO-CEEMDAN and WOA-CEEMDAN. Although VMD achieves a relatively high SNR on both datasets, its RMSE and information entropy remain less favorable than those of the meta-heuristically optimized CEEMDAN variants, indicating that noise suppression alone is insufficient if fault-related information is not preserved effectively.

Overall, Table 5 confirms that introducing meta-heuristic optimization into CEEMDAN is beneficial for cable incipient fault signal decomposition, and that WOA provides the most effective parameter optimization strategy among the methods compared in this study. This result validates the suitability of WOA for adaptively tuning the CEEMDAN parameters (α, n), thereby enhancing the extraction of weak fault characteristics from noisy monitoring signals.

The decomposition level L also plays an important role in CEEMDAN performance. Different decomposition levels affect the number of intrinsic mode functions (IMFs) and the depth of signal decomposition. The results are shown in Figure 12.

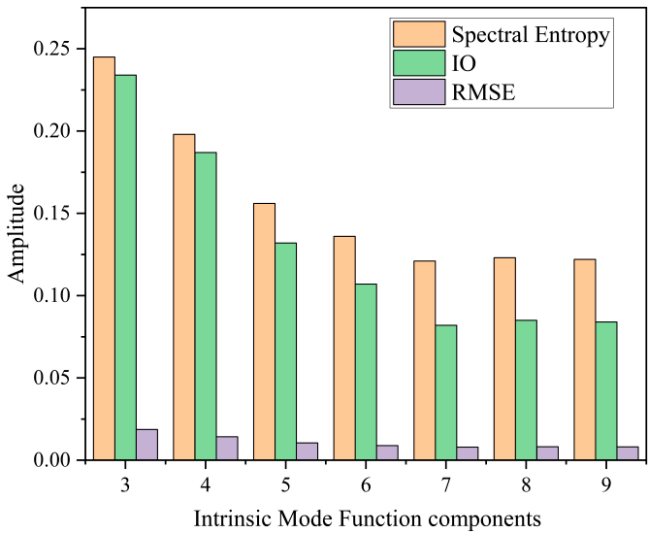


Figure 12. Performance comparison of different diagnostic models on the simulation and real-world datasets.

When the number of IMF layers increases from 3 to 7, the spectral entropy decreases from 0.245 to 0.121 (a reduction of 50.6%), the orthogonality drops from 0.234 to 0.082 (a reduction of 64.9%), and the reconstruction error declines from 0.0187 to 0.0079 (a reduction of 57.7%). This indicates that the first seven IMF layers contain the main characteristic information of early cable fault signals, including high-frequency partial discharge pulses, intermittent arc features, and power frequency fundamental components. When MaxIMF exceeds 7, the improvement in each performance metric becomes extremely limited, while the increased feature dimensionality may interfere with the subsequent classification model. Therefore, this paper selects MaxIMF = 7 as the optimal maximum number of IMF layers.

After determining the optimal decomposition parameters (α = 0.13, n = 126, L = 7), the WOA-CEEMDAN method is applied to the cable incipient fault signals. WOA-CEEMDAN decomposition is performed, as shown in Figure 13.

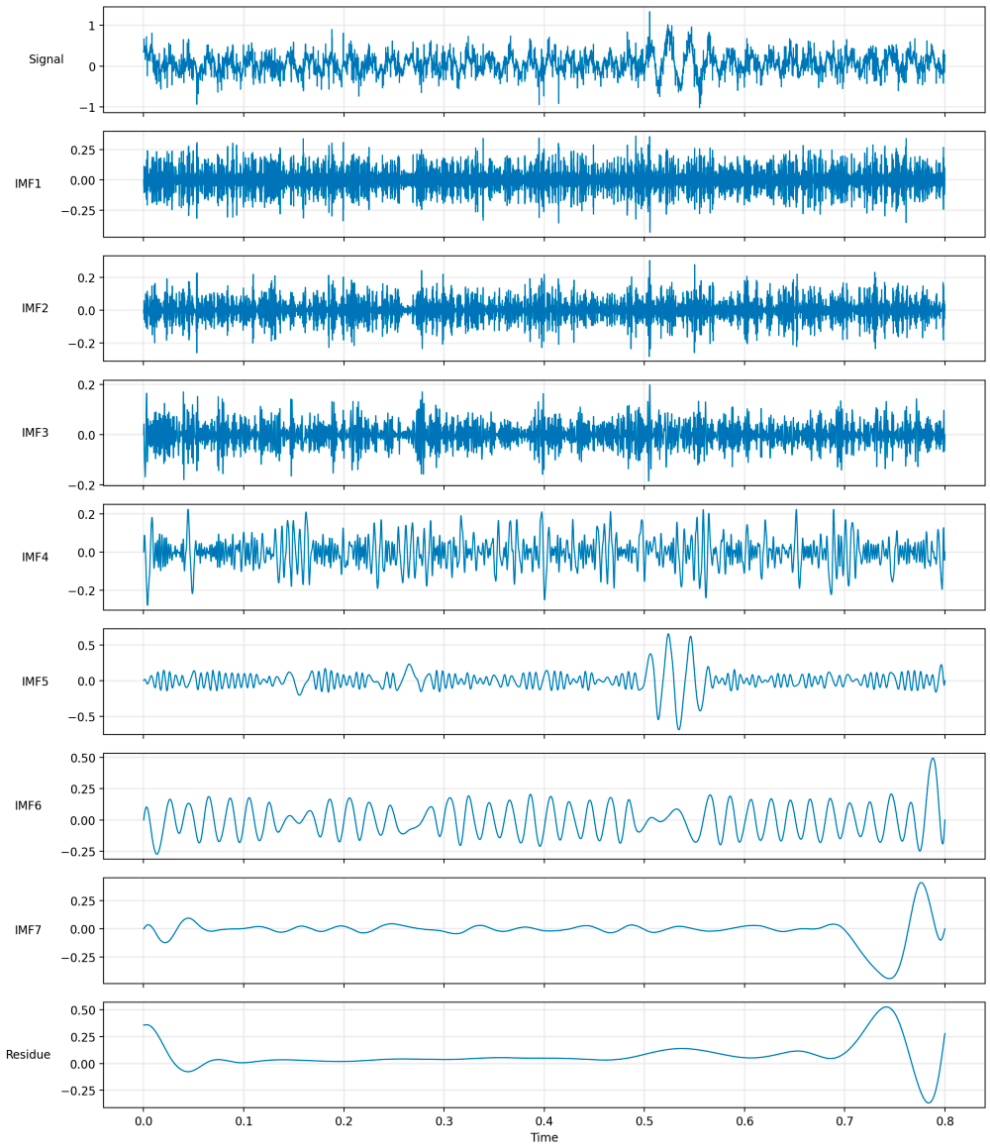


Figure 13. WOA-CEEMDAN decomposition results of a representative incipient-fault signal.

Figure 13 presents the WOA-CEEMDAN decomposition results. IMF1-IMF3 capture high-frequency transient features corresponding to initial fault reflections and subsequent oscillations, serving as primary carriers of early discharge information. IMF4-IMF6 represent medium-to-low frequency components associated with power frequency coupling and load current.

Figure 13 indicate that the decomposed IMFs have clear physical meanings and distinct time–frequency characteristics. IMF1 represents high-frequency random noise (2–5 kHz) with a small amplitude (±0.15 A), high kurtosis (8.2), and an energy contribution of 2.34%. IMF2 is a damped oscillatory mode centered around 800 Hz, corresponding to the electromagnetic transient at arc breakdown, with an amplitude of ±0.35 kA and an energy ratio of 5.67%. IMF3 and IMF4 capture the main fault-related information as regular trapezoidal pulse sequences, associated with the principal arc discharge and its slow modulation, respectively; together they contribute 30.68% of the total energy, with amplitudes up to ±0.65 kA. IMF5 corresponds to the 50 Hz fundamental component and is the dominant energy carrier, accounting for 58.76% of the total energy with an amplitude of ±0.25 kA. IMF6 reflects the low-frequency trend (0–10 Hz) related to the fault-induced DC offset, with an amplitude of ±0.15 kA and an energy contribution of 2.15%.

4.4. Ablation Experiment

To verify the necessity of each component in the proposed hybrid architecture, an ablation study was conducted by removing one module at a time while keeping the data preprocessing procedure, train/validation/test split, CEEMDAN decomposition settings, optimizer, learning rate, batch size, and stopping criteria unchanged. In this way, the contribution of each block can be evaluated under a fair and controlled experimental setting. As shown in Table 6. Four representative variants were considered: (1) TCN only, where only temporal convolutional feature extraction was retained; (2) BiLSTM only, where the model relied only on bidirectional recurrent temporal modeling; (3) TCN-BiLSTM, where the attention module was removed; and (4) the full TCN-BiLSTM-Multihead-Attention model, which corresponds to the proposed framework.

Table 6. Ablation study results of different architectural variants.

Model Variant	Accuracy (%)	Precision (%)	Recall (%)	F1-Score
TCN only	90.36	89.91	89.47	0.897
BiLSTM only	87.92	87.24	87.51	0.874
TCN-BiLSTM	95.08	94.63	94.36	0.945
TCN-BiLSTM-Multihead-Attention	96.21	96.74	96.42	0.965

The purpose of this experiment is to determine whether the performance improvement of the proposed model comes from the combination of all modules rather than from a single dominant block. In particular, TCN is expected to capture local multi-scale transient patterns, BiLSTM is used to model forward and backward temporal dependencies, and the Multihead Attention module is introduced to further emphasize fault-related informative features while suppressing irrelevant fluctuations and residual noise.

The ablation results indicate that the full TCN-BiLSTM-Multihead-Attention model achieves the best overall performance on both datasets. The TCN-only variant performs better than the BiLSTM-only variant, indicating that local multi-scale transient feature extraction is highly important for capturing weak incipient fault signatures. However, without bidirectional temporal modeling, the TCN-only model cannot fully capture long-range sequential dependencies. In contrast, the BiLSTM-only model benefits from temporal context modeling, but its limited local feature extraction capability makes it less effective under noisy and nonstationary signal conditions.

When TCN and BiLSTM are combined, the performance improves substantially, which confirms that local convolutional feature extraction and bidirectional temporal dependency modeling are complementary. After further introducing the Multihead Attention mechanism, all metrics are improved again, indicating that attention-based reweighting helps the model focus on the most fault-sensitive time steps while suppressing irrelevant fluctuations. Therefore, the ablation study verifies that each component contributes positively to the final diagnosis performance, and the superiority of the proposed model arises from the coordinated interaction of all three modules rather than from a single dominant block.

4.5. Comparative Analysis of Different Models

To comprehensively evaluate the effectiveness of the proposed diagnostic framework, two groups of experiments—simulation and real-world—were designed to compare the WOA-CEEMDAN-TCN-BiLSTM-Multihead-Attention model with several representative methods, including Random Forest, 1D-CNN, Bi-LSTM [36], Transformer, and CNN-BiLSTM-Attention [37,38]. The simulation dataset was generated using the PSCAD/EMTDC cable model described in Section 4.1, incorporating controllable fault conditions such as varying fault resistance, inception angles, and discharge magnitudes. The

real-world dataset was collected from an online cable fault acquisition system, containing environmental noise and disturbances that can significantly affect diagnostic performance. Based on these two datasets, each method was comprehensively evaluated in terms of accuracy, precision, recall, and F1-score, and the corresponding results are presented in Table 7.

Table 7. Main parameter settings of the compared models in Table 8.

Model	Main Parameter Settings
Random Forest	Number of trees = 200; maximum depth = 20; minimum samples split = 2; minimum samples leaf = 1; criterion = Gini impurity.
1D-CNN	Three 1D convolutional layers with kernel size = 3; channel numbers = 32, 64, and 64; ReLU activation; global average pooling; fully connected layer with 64 neurons; Softmax output layer.
Bi-LSTM	Two-layer Bi-LSTM; hidden size = 64 in each direction; dropout = 0.3; fully connected layer with 64 neurons; Softmax output layer.
Transformer	Input embedding dimension = 128; number of attention head = 4; number of encoder layers = 2; feed-forward dimension = 256; dropout = 0.3; fully connected layer with 64 neurons; Softmax output layer.
CNN-BiLSTM-Attention	CNN front-end with convolution channels = 32 and 64; kernel size = 3; Bi-LSTM hidden size = 64 in each direction; attention layer with 4 heads; dropout = 0.3; fully connected layer with 64 neurons; Softmax output layer.
WOA-CEEMDAN-TCN-BiLSTM-Multihead-Attention	CEEMDAN search ranges: WOA population size = 30; maximum iterations = 100; spiral constant b = 1.0; optimal CEEMDAN parameters:α = 0.13, n = 126; maximum decomposition level = 7; retained IMFs = IMF1–IMF6; TCN with three residual blocks (kernel size = 3, dilation = 1, 2, 4, channels = 32, 64, 64); Bi-LSTM hidden size = 64 in each direction; Multihead Attention with 4 heads and model dimension = 128;dropout = 0.3;fully connected layer with 64 neurons; Softmax output layer.

Table 8. Overall model comparison on simulation data and real-world data.

Group	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score
Sim data [1]	Random Forest	83.21	82.17	81.74	0.819
	1D-CNN	88.72	87.43	86.98	0.872
	Bi-LSTM	86.25	84.76	85.12	0.849
	Transformer	92.38	91.47	92.05	0.917
	CNN-BiLSTM-Attention	94.21	93.78	93.64	0.937
	WOA-CEEMDAN-TCN-BiLSTM-Multihead-Attention	96.43	97.05	96.88	0.969
Rel data [2]	Random Forest	78.92	77.15	76.42	0.768
	1D-CNN	85.36	84.18	83.72	0.839
	Bi-LSTM	83.21	81.90	82.35	0.821
	Transformer	89.74	88.63	89.05	0.887
	CNN-BiLSTM-Attention	92.56	92.14	91.88	0.920
	WOA-CEEMDAN-TCN-BiLSTM-Multihead-Attention	95.59	96.82	96.50	0.954

[1] Simulation Data in Section 4.1. [2] Real-World Measured Data in Section 4.1.

The proposed WOA-CEEMDAN-TCN-BiLSTM-Multihead-Attention model demonstrates superior and robust performance on both simulation and real datasets, outperforming other methods in accuracy, precision, and F1-score. By integrating WOA-optimized CEEMDAN decomposition with TCN-BiLSTM-Multihead-Attention architecture, the model achieves enhanced feature extraction and global temporal modeling capabilities. While real-world environmental noise causes a slight performance decrease

(96.43%→95.59%) compared to idealized simulation conditions, the model maintains strong practical applicability and generalization capability in realistic scenarios.

4.6. Additional Cross-Cable Validation for Generalizability

To further validate the generalizability of the proposed framework across different cable configurations, additional experiments were conducted on YJV22-8.7/10 kV cables with a length of 8 km and YJV-8.7/10 kV cables with a length of 3 km, in addition to the original YJV42-8.7/10 kV cable used in the main experiments. These supplementary experiments were designed to examine whether the proposed method can maintain stable diagnostic performance under changes in cable structure, cable length, and electrical characteristics. The results demonstrate that the proposed WOA-CEEMDAN-TCN-BiLSTM-Multihead-Attention framework still achieves consistently strong performance on the additional cable models, indicating that the method has satisfactory cross-cable generalizability rather than being restricted to a single experimental setup. As shown in Table 9, the proposed method maintains high accuracy across all three cable types, with only limited performance variation, further supporting its robustness under different cable configurations.

Table 9. Diagnostic performance of the proposed method on different cable types and cable lengths.

Cable Type	Accuracy (%)	Precision (%)	Recall (%)	F1-Score
YJV42-8.7/10 kV	96.85	97.12	96.60	0.968
YJV22-8.7/10 kV	96.12	96.58	95.81	0.962
YJV-8.7/10 kV	95.73	96.05	95.42	0.957

As shown in Table 9, the proposed method maintains high accuracy across all three cable types (95.73–96.85%). This demonstrates strong generalization across cable constructions, attributed to the method proposed in this paper's adaptive optimization that captures fault-related features independent of cable-specific characteristics.

4.7. Small-Sample Experiment

Due to the rarity of incipient fault events in real-world cable monitoring, obtaining large labeled fault datasets is challenging. To evaluate the model's small-sample learning capability, experiments were conducted with four training data proportions (20%, 40%, 60%, and 80%), while keeping the testing dataset unchanged for fair comparison. This approach assesses diagnostic performance under limited data conditions. The results are shown in Table 10.

Table 10. Performance comparison under different training data scales.

Training Data Ratio	1D-CNN	Bi-LSTM	Transformer	C-B-A [1]	OURS [2]
20%	72.84	70.15	78.63	82.47	86.92
40%	79.36	77.42	84.51	88.23	91.65
60%	84.25	82.13	89.74	92.35	94.37
80%	87.92	85.46	92.38	94.21	96.43

[1] CNN-BiLSTM-Attention. [2] WOA-CEEMDAN-TCN-BiLSTM-Multihead-Attention.

The results demonstrate that the proposed WOA-CEEMDAN-TCN-BiLSTM-Multihead-Attention model maintains superior performance even when the available training data are limited. When only 20% of the training data are used, the proposed method still achieves 86.92% accuracy, outperforming the Transformer model by 8.29%. This indicates that the proposed model has strong small-sample learning capability, which is particularly beneficial for early cable fault diagnosis where labeled data are scarce.

4.8. Noise Robustness Experiment Under Gaussian, EMI, and Impulse Noise

To evaluate the model's robustness against real-world disturbances such as electromagnetic interference and environmental noise, Gaussian white noise with varying signal-to-noise ratios (5 dB, 10 dB, 15 dB, and 20 dB) was added to the original signals, testing the proposed method's performance under different noise levels. The results are shown in Table 11.

Table 11. Diagnostic accuracy under Gaussian noise at different SNR levels.

Noise Level (SNR)	1D-CNN	Bi-LSTM	Transformer	C-B-A [1]	OURS [2]
5	64.72	61.85	70.43	75.36	81.54
10	71.94	68.52	78.21	83.41	88.92
15	79.63	76.14	85.32	89.75	93.47
20	84.72	81.36	90.11	93.56	95.84

[1] CNN-BiLSTM-Attention. [2] WOA-CEEMDAN-TCN-BiLSTM-Multihead-Attention.

The results show that the proposed model consistently achieves the highest diagnostic accuracy under all noise levels. Even under severe noise conditions (SNR = 5 dB), the proposed method maintains an accuracy of 81.54%, significantly outperforming the comparison models.

To further evaluate the robustness of the proposed method under practical non-Gaussian interference, additional experiments were conducted under electromagnetic interference (EMI) conditions. In this study, EMI was modeled as sinusoidal narrowband interference superimposed on the original cable current signal. For clarity, three interference severity levels were defined according to the interference amplitude relative to the peak value of the original signal: low EMI corresponds to an interference amplitude of 5% of the signal peak, medium EMI corresponds to 10%, and high EMI corresponds to 15%. The interference frequency was selected from typical power-related electromagnetic components, including 50 Hz, 150 Hz, and 250 Hz, in order to simulate realistic electromagnetic coupling in cable-monitoring environments. The corresponding robustness results under EMI conditions are shown in Figure 14.

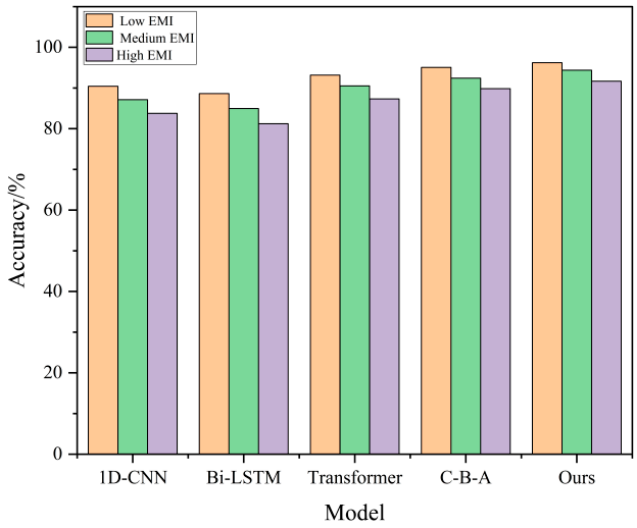


Figure 14. Diagnostic performance of different models under EMI conditions.

As shown in Figure 14, all compared models exhibit a gradual decrease in diagnostic accuracy with increasing EMI intensity. The proposed WOA-CEEMDAN-TCN-BiLSTM-Multihead-Attention framework achieves accuracies of 96.21%, 94.37%, and 91.68% under low-, medium-, and high-level EMI, respectively, which are consistently higher than those

of CNN-BiLSTM-Attention (95.06%, 92.41%, and 89.85%), Transformer (93.18%, 90.52%, and 87.33%), 1D-CNN (90.42%, 87.15%, and 83.76%), and Bi-LSTM (88.63%, 84.97%, and 81.24%). This demonstrates that the proposed method has stronger robustness against narrowband electromagnetic interference.

In addition to EMI, impulse noise was further introduced to simulate abrupt transient disturbances that may arise from switching operations, external electrical shocks, or sensor spikes in industrial environments. The impulse noise was modeled as sparse high-amplitude perturbations randomly occurring in the signal sequence. The disturbance severity was classified into three levels using both impulse probability and impulse amplitude. Specifically, low impulse noise corresponds to an impulse occurrence probability of 0.5% with an amplitude of 2 times the original signal peak, medium impulse noise corresponds to a probability of 1% with an amplitude of 3 times the signal peak, and high impulse noise corresponds to a probability of 2% with an amplitude of 5 times the signal peak. The robustness results under these impulse-noise conditions are presented in Figure 15.

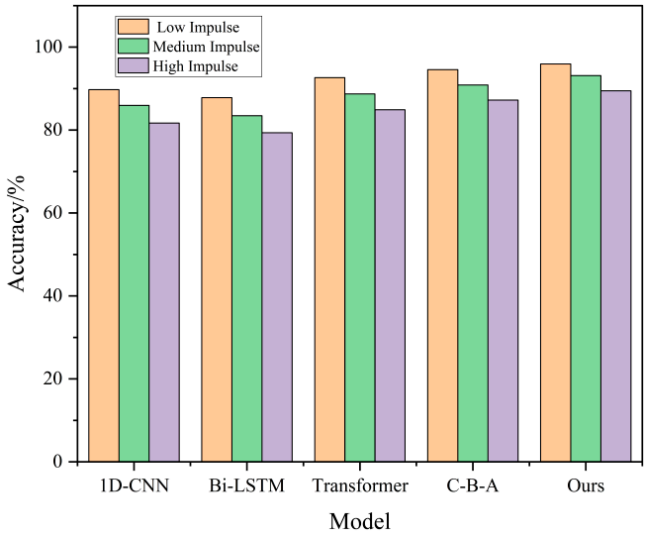


Figure 15. Diagnostic performance of different models under impulse-noise conditions.

As illustrated in Figure 15, all compared models exhibit a clear reduction in diagnostic accuracy as the severity of impulse noise increases, reflecting the strong influence of sparse high-amplitude transients on fault-feature extraction and classification. The proposed WOA-CEEMDAN-TCN-BiLSTM-Multihead-Attention framework achieves accuracies of 95.93%, 93.14%, and 89.47% under low-, medium-, and high-level impulse-noise conditions, respectively, which are consistently higher than those of CNN-BiLSTM-Attention (94.58%, 90.86%, and 87.24%), Transformer (92.64%, 88.73%, and 84.91%), 1D-CNN (89.75%, 85.94%, and 81.68%), and Bi-LSTM (87.82%, 83.46%, and 79.35%). This confirms that the proposed method maintains superior robustness under abrupt nonstationary interference.

4.9. ROC Curve and AUC Analysis

Receiver Operating Characteristic (ROC) curves were used to evaluate the classification performance of different models. The ROC curve illustrates the relationship between the True Positive Rate (TPR) and False Positive Rate (FPR) under varying decision thresholds. The Area Under the Curve (AUC) quantitatively reflects the classifier's discrimination ability.

Figure 16 shows the ROC curves of several representative models, including 1D-CNN, Bi-LSTM, Transformer, and the proposed WOA-CEEMDAN-TCN-BiLSTM-Multihead-Attention model. The corresponding AUC values are summarized in Table 12.

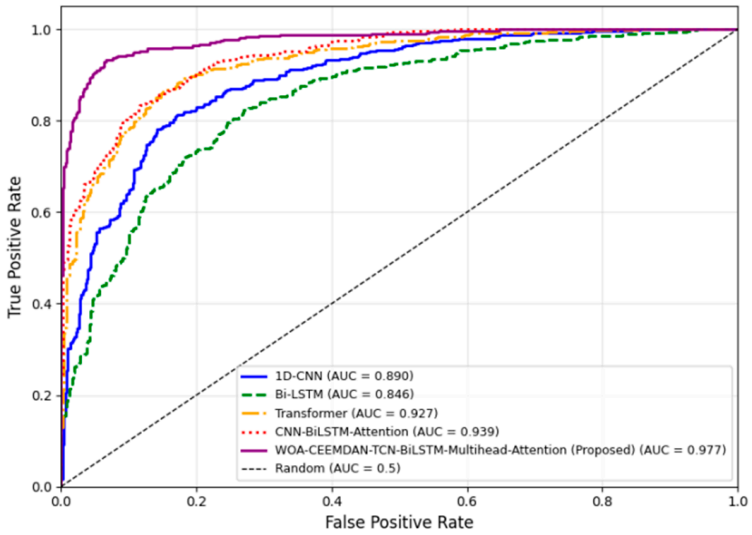


Figure 16. ROC curves of different diagnostic models on the simulation and real-world datasets.

Table 12. AUC comparison of different models on the simulation and real-world datasets.

Model	Sim AUC [1]	Rel AUC [2]
1D-CNN	0.88	0.89
Bi-LSTM	0.86	0.84
Transformer	0.92	0.92
CNN-BiLSTM-Attention	0.94	0.93
OURS [3]	0.97	0.97

[1] Simulation Data in Section 4.1. [2] Real-World Measured Data in Section 4.2. [3] WOA-CEEMDAN-TCN-BiLSTM-Multihead-Attention.

Figure 16 shows that the ROC curve of the proposed WOA-CEEMDAN-TCN-BiLSTM-Multihead-Attention model remains consistently closer to the upper-left corner on both the simulation and real-world datasets, indicating a more favorable balance between fault sensitivity and false alarm rate across a wide range of thresholds. The corresponding AUC values are summarized in Table 12. On the simulation dataset, the proposed model achieves the highest AUC of 0.97, improving upon CNN-BiLSTM-Attention by 0.03, Transformer by 0.05, 1D-CNN by 0.09, and Bi-LSTM by 0.11. A similar trend is observed on the real-world dataset, where the proposed model also attains an AUC of 0.97, which is 0.04 higher than CNN-BiLSTM-Attention and 0.05 higher than Transformer.

These results indicate that the proposed framework does not merely perform well at a single operating point; rather, it preserves strong separability between incipient-fault and non-fault patterns over a broad range of decision thresholds. This property is particularly important for engineering deployment, because alarm thresholds in online monitoring systems often need to be adjusted according to different operational risk preferences. The high AUC on the real-world dataset further confirms that the combination of WOA-CEEMDAN preprocessing and hybrid sequence modeling effectively suppresses the influence of noise and distribution variability.

It is also worth noting that the AUC gap between the simulation and real-world datasets is negligible for the proposed method, whereas several comparison models exhibit a more obvious performance drop under real measurement conditions. This observation

suggests that the proposed framework has better cross-scenario generalization and more stable probability discrimination capability. Therefore, the ROC/AUC analysis, together with the accuracy, F1-score, and confusion-matrix results, provides consistent evidence that the proposed method offers superior diagnostic reliability for practical incipient cable fault monitoring.

4.10. Model Performance Analysis

To more clearly illustrate the recognition performance of the proposed WOA-CEEMDAN-TCN-BiLSTM-Multihead-Attention model on real-world collected data, this section presents its confusion matrix as shown in Figure 17. On the real-world test set, the model achieves 1380 true positives (TP) and 636 true negatives (TN), along with 39 false positives (FP) and 45 false negatives (FN). This indicates that the model exhibits a slight preference for issuing conservative alarms rather than missing actual faults—a desirable characteristic for early warning applications in power systems. Accordingly, the false negative rate (FNR) is approximately 3.16%, while the false positive rate (FPR) stands at about 5.78%. In other words, the proposed method maintains a high probability of detecting incipient cable faults before they evolve into more severe failures. The low FNR is important for early fault diagnosis, since missed alarms may allow weak insulation defects to develop into more severe failures. Meanwhile, the relatively limited FPR suggests that the framework also has good practicality for online monitoring scenarios, where excessive false alarms may increase maintenance burden.

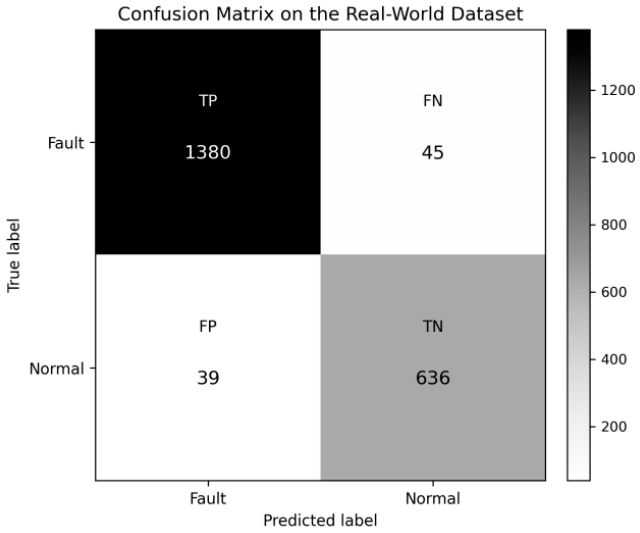


Figure 17. Confusion matrix of the proposed model on the real-world test dataset.

Overall, the confusion-matrix-based analysis further confirms that the proposed model is not only accurate in terms of global metrics, but also practically meaningful in engineering applications where fault sensitivity is more critical than nominal-state conservatism.

In addition, the Matthews correlation coefficient (MCC) calculated from the confusion matrix is 0.909, which further confirms the balanced classification capability of the proposed model. Unlike accuracy alone, MCC jointly considers true positives, true negatives, false positives, and false negatives, and therefore provides a more comprehensive evaluation of diagnostic reliability. Overall, the confusion-matrix-based analysis demonstrates that the proposed framework achieves strong and balanced performance on real-world cable fault data.

To further assess the balance between precision and recall for different diagnostic models, the Precision–Recall AUCs are shown in Figure 18.

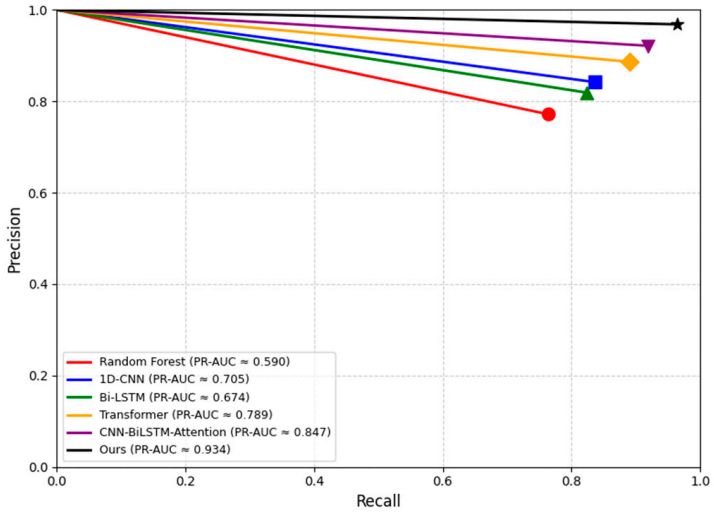


Figure 18. Precision–recall curves of different diagnostic models (Blue represents WOA-CEEMDAN preprocessing, purple denotes the hybrid learning model, and green indicates the output).

As shown in Figure 18, the proposed model achieves a more favorable precision–recall trade-off than the comparison methods. Its PR curve remains closer to the upper-right region, indicating that high fault recall can still be maintained without a substantial loss of precision. This result is consistent with the previously reported accuracy, F1-score, ROC-AUC, and confusion-matrix analysis, and further confirms the superior reliability of the proposed framework for practical incipient cable fault diagnosis.

5. Conclusions

This study developed an incipient fault diagnosis framework for power cables by combining WOA-optimized CEEMDAN with a TCN-BiLSTM-Multi-HeadAttention classifier. By mapping WOA directly to the CEEMDAN parameter-optimization problem, the proposed method improved decomposition quality and enhanced the extraction of weak transient fault features from noisy signals. The optimized decomposition achieved its best performance at $\alpha = 0.13$ and $n = 126$, with an RMSE of 0.097 and an SNR of 8.42 dB. On the real-world test dataset, the proposed model achieved 96.00% accuracy, 97.25% precision, 96.84% recall, an F1-score of 0.970, and an AUC of 0.97, confirming strong and balanced diagnostic performance. Additional small-sample and noise-robustness experiments further showed that the proposed framework remains effective under limited training data and strong interference. Overall, the results indicate that the proposed method provides an effective solution for early cable-fault diagnosis.

6. Limitations and Future Work

This study has several limitations. First, although both simulation and measured data were used, the real-world data were collected under a relatively limited acquisition framework; therefore, the proposed method has not yet been fully validated on independent cross-dataset or cross-domain scenarios. Second, the PSCAD model provides a controllable simulation environment but cannot fully reproduce all practical cable operating conditions, such as complex aging processes, environmental variability, and sensor-related disturbances. Third, the present framework was validated as an offline diagnosis method. In future work, an online monitoring system will be implemented on hardware platforms to further evaluate the real-time applicability and engineering feasibility of the proposed method.

Author Contributions: Conceptualization and methodology, Y.X. and Y.Y.; software, Y.Y.; validation, Y.Y.; formal analysis, Y.X. and Y.Y.; investigation, Y.Y.; resources, Y.Y.; data curation, Y.Y.; writing—original draft preparation, Y.Y.; writing—review and editing, Y.Y.; visualization, Y.Y.; supervision, Y.X. and Y.Y.; project administration, Y.Y.; funding acquisition, Y.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

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

References

1. Sidhu, T.S.; Xu, Z. Detection of incipient faults in distribution underground cables. *IEEE Trans. Power Deliv.* **2010**, *25*, 1363–1371. [CrossRef]
2. Dong, X.; Yang, Y.; Zhou, C.; Hepburn, D.M. Online Monitoring and Diagnosis of HV Cable Faults by Sheath System Currents. *IEEE Trans. Power Deliv.* **2017**, *32*, 2281–2290. [CrossRef]
3. Mirzaei, M.; Ab Kadir, M.Z.; Moazami, E.; Hizam, H. Review of fault location methods for distribution power system. *Aust. J. Basic Appl. Sci.* **2009**, *3*, 2670–2676.
4. Peng, N.; Zhang, Z.; Liang, R.; Jiang, C.; Zhang, P.; Ren, X.; Wang, X. Fault sensing of the distribution cable feeders by time-domain measurements. *IEEE Trans. Ind. Inform.* **2022**, *19*, 8170–8182. [CrossRef]
5. Haohao, C.; Jing, L.; Ping, C. Design of fault location algorithm based on online distributed travelling wave for HV power cable. *PLoS ONE* **2024**, *19*, e0296513. [CrossRef] [PubMed]
6. Peng, N.; Zhang, W.; Liang, R.; Du, M.; Zhang, P.; Wang, W.; Wang, W.; Hu, Y. Fault Distance Estimation for High-Voltage Cables Down the Mines Based on Voltage Continuity Considering Distributed Parameters. *IEEE Trans. Power Deliv.* **2025**, *40*, 1811–1824. [CrossRef]
7. Peng, N.; Liu, Z.; Xie, Z.; Liang, R.; Zhang, P.; Wang, W.; Li, Y.; Fan, W.; Du, W.; Zhu, S. High-impedance Fault Detection and Location Methods for Three-core Submarine Cables Based on Voltage Distribution Features. In *IEEE Transactions on Power Delivery*; IEEE: New York, NY, USA, 2026. [CrossRef]
8. Mallat, S.G. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *11*, 674–693. [CrossRef]
9. Liu, X.; Shi, G.; Liu, W. An improved empirical mode decomposition method for vibration signal. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 5525270. [CrossRef]
10. Chen, W.; Xiong, C.; Yu, L.; Lian, S.; Ye, Z. Dynamic monitoring of an offshore jacket platform based on RTK-GNSS measurement by CF-CEEMDAN method. *Appl. Ocean Res.* **2021**, *115*, 102844. [CrossRef]
11. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2013**, *62*, 531–544. [CrossRef]
12. El Din, E.T.; Gilany, M.; Aziz, M.A.; Ibrahim, D.K. A wavelet-based fault location technique for aged power cables. In *Proceedings of the IEEE Power Engineering Society General Meeting, San Francisco, CA, USA, 16 June 2005*; IEEE: Piscataway, NJ, USA, 2005; pp. 2485–2491.
13. De Mello, R.F.; Ponti, M.A. Statistical learning theory. In *Machine Learning: A Practical Approach on the Statistical Learning Theory*; Springer: Berlin/Heidelberg, Germany, 2018; Volume 3.
14. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
15. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
16. Yan, J.; Zhang, Y.; Su, Q.; Li, R.; Li, H.; Lu, Z.; Lu, H.; Lu, Q. Time series prediction based on LSTM neural network for top tension response of umbilical cables. *Mar. Struct.* **2023**, *91*, 103448. [CrossRef]
17. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017.
18. Wang, F.; Zhang, P.; Li, J.; Li, Z.; Zhao, M.; Liang, Y.; Su, G.; You, X. Multi-feature based extreme learning machine identification model of incipient cable faults. *Front. Energy Res.* **2024**, *12*, 1364528. [CrossRef]
19. Fu, H.; Qiu, L.; Ai, Y.; Tu, J.; Yan, Y. Deep learning-based fault detection and location in underground power cables using resonance frequency analysis. *Electr. Eng.* **2025**, *107*, 4051–4062. [CrossRef]
20. Shao, Q.; Fan, S.; Zhang, Z.; Liu, F.; Fu, Z.; Lv, P.; Mu, Z. Artificial intelligence in cable fault detection and localization: Recent advances and research challenges. *Energies* **2025**, *18*, 3662. [CrossRef]

21. Ali, Z.M.; Esmail, E.M. Deep learning and wavelet packet transform for fault diagnosis in double circuit transmission lines. *Sci. Rep.* **2025**, *15*, 30145. [CrossRef] [PubMed]
22. Chi, P.; Liang, R.; Hao, C.; Li, G.; Xin, M. Cable fault diagnosis with generalization capability using incremental learning and deep convolutional neural network. *Electr. Power Syst. Res.* **2025**, *241*, 111304. [CrossRef]
23. He, J.; Zhao, H. Fault diagnosis and location based on graph neural network in telecom networks. In *Proceedings of the 2020 International Conference on Networking and Network Applications (NaNA), Haikou City, China, 10–13 December 2020*; IEEE: Piscataway, NJ, USA, 2020.
24. Zhao, Y.; Xiong, Y.; Guo, T.; Shang, Y.; Wang, W. Research on cable fault detection algorithm based on improved neural network algorithm. In *Proceedings of the 2024 IEEE 4th International Conference on Electronic Technology, Communication and Information (ICETCI), Changchun, China, 24–26 May 2024*; IEEE: Piscataway, NJ, USA, 2024.
25. Laganà, F.; Pratticò, D.; Quattrone, M.F.; Pullano, S.A.; Calcagno, S. Hybrid AI–Taguchi–ANOVA Approach for Thermographic Monitoring of Electronic Devices. *Eng* **2026**, *7*, 28. [CrossRef]
26. Angiulli, G.; Versaci, M.; Burrascano, P.; Laganà, F. A Data-Driven Gaussian Process Regression Model for Concrete Complex Dielectric Permittivity Characterization. *Sensors* **2025**, *25*, 6350. [CrossRef]
27. Pratticò, D.; Laganà, F.; Oliva, G.; Fiorillo, A.S.; Pullano, S.A.; Calcagno, S.; De Carlo, D.; La Foresta, F. Sensors and Integrated Electronic Circuits for Monitoring Machinery on Wastewater Treatment: Artificial Intelligence Approach. In Proceedings of the 2024 IEEE Sensors Applications Symposium (SAS), Naples, Italy, 23–25 July 2024; pp. 1–6. [CrossRef]
28. Mirjalili, S.; Lewis, A. The whale optimization algorithm. *Adv. Eng. Softw.* **2016**, *95*, 51–67. [CrossRef]
29. Chen, X.; Cheng, L.; Liu, C.; Liu, Q.; Liu, J.; Mao, Y.; Murphy, J. A WOA-based optimization approach for task scheduling in cloud computing systems. *IEEE Syst. J.* **2020**, *14*, 3117–3128. [CrossRef]
30. Nasiri, J.; Khiyabani, F.M. A whale optimization algorithm (WOA) approach for clustering. *Cogent Math. Stat.* **2018**, *5*, 1483565. [CrossRef]
31. Thu, N.T.H.; Bao, P.Q.; Nam, N.V.N. Multiple Step Ahead Forecasting of Rooftop Solar Power Based on a Novel Hybrid Model of CEEMDAN-Bidirectional LSTM Network with Structure Optimized by PSO Method. In Proceedings of the 11th International Conference on Control, Automation and Information Sciences (ICCAIS), IEEE, Hanoi, Vietnam, 21–24 November 2022; pp. 522–528.
32. Bai, S.; Kolter, J.Z.; Koltun, V. Convolutional sequence modeling revisited. In Proceedings of the 6th International Conference on Learning Representations, ICLR, 2018, Vancouver, BC, Canada, 30 April–3 May 2018.
33. Li, Q.; Luo, H.; Cheng, H.; Deng, Y.; Sun, W.; Li, W.; Liu, Z. Incipient fault detection in power distribution system: A time–frequency embedded deep-learning-based approach. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–14. [CrossRef]
34. Li, X.; Zhang, W.; Ding, Q. Understanding and improving deep learning-based rolling bearing fault diagnosis with attention mechanism. *Signal Process.* **2019**, *161*, 136–154. [CrossRef]
35. Zhang, R.; Wang, Y.; Wan, X.; Ming, Y.; Yang, S. Position prediction of underwater gliders based on a new heterogeneous model ensemble method. *Ocean Eng.* **2024**, *309*, 118312. [CrossRef]
36. Siami-Namini, S.; Tavakoli, N.; Namin, A.S. The performance of LSTM and BiLSTM in forecasting time series. In Proceedings of the IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 9–12 December 2019; pp. 3285–3292.
37. Bu, Q.; Lyu, P.; Sun, R.; Jing, J.; Lyu, Z.; Hou, S. Fault diagnosis method using CNN-attention-LSTM for AC/DC microgrid. *Modelling* **2025**, *6*, 107. [CrossRef]
38. Zhang, J.; Ye, L.; Lai, Y. Stock price prediction using CNN-BiLSTM-Attention model. *Mathematics* **2023**, *11*, 1985. [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.