EAI Endorsed Transactions
on Energy Web
Research Article EAI.EU

Meteorological Factor-Based Early Warning of Power Grid Faults via the Meta-Fuse Framework

Yong Jiang[1], Yuanqing Shi[1], Weifeng Cao[1], Renzhi Zhang[2,*], Zehui Xie[1], Wei Deng[1]

[1]College of Electric and Information Engineering, Zhengzhou University of Light Industry, Dong Feng Road, Zhengzhou, 450002, Henan Province, Peoples Republic of China.
[2]College of Information Engineering, Huanghuai University, 76 Kaiyuan Avenue, Zhumadian, 463000, Henan Province, Peoples Republic of China.

Abstract

INTRODUCTION: Extreme weather events increasingly threaten power grid resilience, causing cascading failures, while conventional models fail to capture complex interactions among meteorological, topological, and spatiotemporal factors.
OBJECTIVES: This paper aims to develop Meta-Fuse, a meteorological attention-based fused model, for early warning of power grid faults under environmental uncertainty.
METHODS: This study utilizes methods grounded in the analysis of social networks. The proposed framework integrates Triangular Topology Aggregation Optimization, a CNN-BiLSTM pipeline for spatiotemporal feature extraction, and a multi-head attention classifier, fused via a hybrid CRITIC-entropy strategy.
RESULTS: The model achieves a classification accuracy of 0.9464 with incomplete weather data and maintains a sensitivity of 0.9309 with high-dimensional features, outperforming traditional architectures.
CONCLUSION: Meta-Fuse demonstrates effective topology-informed learning and multimodal fusion, confirming its value for proactive resilience strategies in weather-exposed power systems.

Keywords: Power grid fault prediction, Convolutional Neural Networks, Multi-Head Attention, BiLSTM.

Received on 31 December 2025, accepted on 31 March 2026, published on 14 April 2026

doi: 10.4108/ew.11474

*Corresponding author. Email: 20070784@huanghuai.edu.cn

1. Introduction

Extreme weather events—such as storms, heatwaves, and torrential rainfall—are becoming increasingly frequent and severe, posing a major risk to the stability and functionality of modern power grids [1]. This challenge is particularly acute in regions like China. Rapid grid expansion across diverse terrains has increased the infrastructure's exposure to meteorological stressors, which often trigger cascading failures, widespread service interruptions, and significant safety risks [2].Furthermore, the integration of renewable resources adds complexity to maintaining transient stability under such compound disturbances [3]. While traditional fault analysis cites equipment aging, operational errors, and environmental volatility as key contributors[4–5], the non-stationary and spatially correlated nature of weather-induced faults demands predictive models that go beyond treating weather as a mere input variable. The core research gap lies in developing frameworks that can dynamically model the interplay between evolving grid topology under physical stress, multi-modal spatiotemporal data, and the imbalanced nature of fault events during extremes.

Current research leveraging deep learning for grid resilience can be categorized into three complementary yet

distinct strands, each with limitations when applied to meteorological fault prediction:

(1).Topology-Aware and Geometric Deep Learning: A notable methodological development utilizes graph neural networks (GNNs) to represent electrical grid interconnectivity for anticipating outages. Illustratively, Varbella et al. applied geometric deep learning to forecast cascading failures in real-time, thereby directly leveraging the system's topological configuration [1]. However, these approaches typically treat the topology as static. They do not model how the effective electrical or logical topology deforms under meteorological stress (e.g., line tripping due to high winds, dynamic re-routing), which is crucial for accurate pre-fault anticipation rather than post-initiation cascade tracking.

(2).Spatiotemporal and Multi-Modal Feature Learning: The integration of temporal and spatial features is recognized as vital. Methods combining Temporal Convolutional Networks (TCN) with Bidirectional Long Short-Term Memory (BiLSTM) networks have shown efficacy in processing time-series data for tasks like state estimation[6]. Furthermore, the fusion of multi-modal data (e.g., electrical signals, text, images) is an active frontier. A recent study by Jiahao Zhang and colleagues introduced a semi-supervised learning model featuring an adaptive threshold mechanism, designed specifically for identifying faults in urban power distribution networks [7]. In contrast, a separate 2024 study employed a cross-modal attention fusion (CMAF) mechanism, designed to combine complementary features extracted from heterogeneous data sources for enhanced fault identification [8]. Despite their sophistication, these models often lack a dedicated mechanism to weight meteorological features adaptively or to adjust their sensitivity based on real-time weather severity, leading to suboptimal performance during rare but critical extreme events.

(3).Meta-Heuristic Optimization for Power Systems: Optimization algorithms are widely used for tuning model parameters and solving complex grid problems. The Triangular Topology Aggregation Optimizer (TTAO) introduced by Zhao et al.[9] offers a topology-informed strategy for model tuning. Similarly, metaheuristics like Atom Search Optimization (ASO) have been applied to problems such as fault location in transmission lines[10]. A common limitation is their assumption of a static optimization landscape or topological configuration, making them sensitive to the non-stationary conditions induced by varying weather [11]. Moreover, they are not inherently designed to handle the severe class imbalance where fault samples during major storms are vastly outnumbered by normal operation data. Recent analyses of large-scale power outages further underscore the critical impact of compounding meteorological stressors on grid resilience [12].

To address these interconnected gaps, this study proposes Meta-Fuse—a Meteorological Attention-based Fused model. Meta-Fuse introduces three integrated innovations designed to specifically overcome the limitations outlined above: (1) A storm-adaptive topological deformation module that extends static TTAO principles to model grid reconfiguration under meteorological stress; (2) A meteorology-aware multimodal fusion mechanism that uses a modularized attention architecture comprising multiple parallel attention mechanisms is integrated to dynamically weigh the importance of CNN-extracted spatial features and BiLSTM-captured temporal dynamics based on concurrent weather perturbation inputs; and (3) An adaptive anomaly thresholding scheme that aligns classification sensitivity with real-time meteorological severity indices, directly tackling the class imbalance problem during extreme events.

Built upon a hybrid architecture that cohesively integrates a weather-adapted TTAO, a CNN–BiLSTM backbone, and a meteorological attention layer, the Meta-Fuse framework is designed to explicitly capture the complex, dynamic interplay between a stress-deformable grid topology, multi-modal environmental conditions, and incipient fault propagation. Validated on historical fault logs synchronized with real-time weather data, our method demonstrates superior performance in delivering accurate, stable, and early fault predictions, meeting the heightened reliability and timeliness demands of modern power systems operating in an era of climate change.

2. Fundamental Principles

2.1 Triangular Topology Aggregation Optimizer (TTAO)

The TTAO algorithm is a meta-heuristic optimizer based on the geometric evolution of similar triangles. It consists of three phases: initialization, dynamic topological reconstruction, and cooperative aggregation, each mathematically formulated as follows[13].

(1). Population Initialization:

For a population of size N in D-dimensional space, N/3 triangular units are formed. The position of the first vertex in the i-th unit is initialized as:

$$\overrightarrow{x_{i1}} = r_0 \times \left(\overrightarrow{UB} - \overrightarrow{LB}\right) + \overrightarrow{LB} \tag{1}$$

where $r_0 \in [0,1]$ denotes a randomly generated value., and $\overrightarrow{LB}$, $\overrightarrow{UB}$ mark the permissible range..

(2). Dynamic Topological Reconstruction:

The second and third vertices are generated via directional vectors derived from spherical coordinates:

$$\overrightarrow{X_{I2}} = \overrightarrow{X_{i2}} + l \times \overrightarrow{f(\theta)} \tag{2}$$

$$\overrightarrow{X_{i3}} = \overrightarrow{X_{i1}} + l \times \overrightarrow{f\left(\theta + \frac{\pi}{3}\right)} \tag{3}$$

where $l = 9 \cdot e^{-t/T}$ controls the triangle size, t corresponds to the ongoing iteration, T is the maximum iterations, and:

$$\overrightarrow{f(\theta)} = [\cos\theta_1, \cos\theta_2, \ldots, \cos\theta_{D-1}, \cos\theta_D] \quad (4)$$

(3). Internal Aggregation:

A fourth vertex is formed inside each triangle by linear combination:

$$\overrightarrow{X_{i4}} = r_1 \overrightarrow{X_{i1}} + r_2 \overrightarrow{X_{i2}} + r_3 \overrightarrow{X_{i3}} \quad (5)$$

with $r_1, r_2, r_3 \in [0,1]$ and $r_1 + r_2 + r_3 = 1$.

(4). Global Aggregation:

Information is exchanged between the best vertex of unit i and a randomly selected unit:

$$\overrightarrow{X_{i,new1}^{t+1}} = r_4 \times \overrightarrow{X_{i,best}^{t}} + (1 - r_4) \times \overrightarrow{X_{tand,best}^{t}} \quad (6)$$

A greedy update is applied if the new solution improves fitness.

(5). Local Aggregation:

To avoid local optima, a local search is performed:

$$\overrightarrow{X_{i,new2}^{t+1}} = \overrightarrow{X_{i,best}^{t+1}} + a\left(\overrightarrow{X_{i,best}^{t+1}} - \overrightarrow{X_{i,sbest}^{t+1}}\right) \quad (7)$$

where $a = \ln\left(\frac{e - e^3}{T - 1} \cdot \mathrm{t} + e^3\right)$.

These formulations enable TTAO to dynamically explore and exploit the search space, thereby enhancing convergence in hyperparameter optimization for fault prediction.

2.2 Convolutional Neural Network (CNN)

A complete CNN architecture is commonly composed of a hierarchical arrangement of specialized layers. These generally include an input layer, followed by successive stages of convolutional operations, non-linear activation functions, and down-sampling through pooling. The network ultimately culminates in one or more fully connected layers leading to a final output layer.

Input Normalization: Normalize input data (e.g., using Z-score normalization) to accelerate training and enhance model stability. Convolutional Layer: Employ convolutional kernels (filters) that slide over the input, computing the weighted sum of local regions as follows:

$$T_i^k = f\left\{\sum_{j \in M_i} T_i^k \times m_{ij}^k + b_i^k\right\} \quad (8)$$

Where, T_i^k is the i-th feature region of layer K, m_{ij}^k represents the **j**-th parameter of the **i**-th convolutional filter in layer **K**., M_i is the current feature input set, b_i^k is the network offset top, and F is the activation parameter.

2.3 Bidirectional Long Short-Term Memory Network (BiLSTM)

BiLSTM is a specialized variant of Long Short-Term Memory (LSTM) networks that propagates information in both forward and backward directions, effectively capturing the interdependencies between past and future data points in a time series, thereby enhancing the extraction of temporal features[14].

Within a BiLSTM architecture, the input sequence is concurrently propagated through two independent temporal processing pathways. Each directional layer maintains its own hidden state representations, and the resultant output is synthesized by integrating the information from both of these layers. The specific computation is as follows:

$$\overrightarrow{h_t} = f\left(\overrightarrow{w_x} \cdot X_t + \overrightarrow{h_{t-1}} \cdot \overrightarrow{w_h} + \overrightarrow{b_h}\right) \quad (9)$$

$$\overleftarrow{h_t} = f\left(\overleftarrow{w_x} \cdot X_t + \overleftarrow{h_{t-1}} \cdot \overleftarrow{w_h} + \overleftarrow{b_h}\right) \quad (10)$$

$$Y_t = f\left(\overrightarrow{w_y} \cdot \overrightarrow{h_t} + \overleftarrow{h_t} \cdot \overrightarrow{w_y} + b_y\right) \quad (11)$$

Here, f denotes the non-linear activation function, while w and b represent the trainable weight and bias parameters, respectively. The result generated by the forward layer is $\overrightarrow{h_t}$, and the corresponding result from the backward layer is $\overleftarrow{h_t}$ .These two vectors are subsequently integrated to produce the final output, Y_t.

These two vectors are subsequently integrated to produce the final output, Y.

2.4 Multi-Head Attention Mechanism

The multi-head attention mechanism extracts correlation information along the temporal dimension of data. Compared to recurrent neural networks, self-attention mechanisms can better handle the long-term dependency issues inherent in RNNs[15], offering a powerful alternative for sequence modeling as demonstrated in foundational work [16].

In the attention mechanism, the input data $I \in R^{n \times T}$ is processed through linear layers to compute three intermediate matrices: Q、K、$V \in R^{n \times T}$:

$$A = Softmax\left(\frac{mask\left(K^T \times Q\right)}{\sqrt{n}}\right) \quad (12)$$

$$O = A \times V \quad (13)$$

Q and K are used to calculate the attention matrix $A \in R^{T \times T}$, which represents the degree of correlation between elements within the sequence. The mask operation modifies the attention matrix by setting certain elements to zero, ensuring that each time step attends only to its preceding steps and not to future ones, this architectural design effectively blocks the flow of information from subsequent time steps. The final output of the attention

module is obtained by multiplying the attention matrix $O = \in R^{n \times T}$.

The multi-head attention mechanism improves the model's capacity to extract diverse dependencies in the data by utilizing several parallel attention heads. This process involves aggregating the outputs from all heads and transforming them linearly, as defined in the following equation:

$$MHA(I) = W^{O} \times \left[A_1(I), \ldots, A_m(I) \right] \quad (14)$$

In this notation, head_h refers to the result generated by the h-th attention module, while WO is the corresponding output projection matrix. These components work in concert to allow the model to identify intricate relationships and dependencies within the input data, such as long-range contextual and syntactic patterns, ultimately improving its overall accuracy and generalization capability.

3. Model Architecture

3.1 Basic Structure of the Meta-Fuse Framework

To address the challenges of spatiotemporal complexity and meteorological uncertainty in power grid fault diagnosis, this study constructs a power grid meteorological fault diagnosis model based on CNN-BiLSTM-MHA, integrating CNN, BiLSTM, and Multi-Head Attention (MHA). The overall architecture is graphically summarized in Figure 1.

The input to the framework consists of preprocessed fault-related time series data. First, The CNN module employs convolutional filters that traverse the input data to capture local spatial patterns, effectively reducing noise and preserving location-aware features. The extracted topographical characteristics are subsequently processed by the bidirectional LSTM, which models both forward and backward sequential interdependencies, thereby encapsulating comprehensive contextual information across the temporal dimension, enabling the model to understand the evolution of fault signatures over time. The synergistic combination of CNN and BiLSTM has proven effective for temporal pattern recognition in various mechanical and electrical fault diagnosis tasks [17-19].

In the specific implementation, the CNN module consists of two convolutional layers with kernel size 3×1, containing 16 and 32 feature maps, respectively, each followed by a ReLU activation function. After feature extraction, the output is flattened and fed into a BiLSTM layer, whose number of hidden units is adaptively determined by the TTAO optimizer. Subsequently, a multi-head self-attention layer is applied to the BiLSTM outputs to dynamically reweight temporally salient features before final classification.

To further enhance feature discrimination, a Multi-Head Attention (MHA) mechanism is introduced. This module dynamically assigns attention weights to the outputs of the BiLSTM layer. This capability allows the model to prioritize the most salient temporal intervals while downplaying non-essential or extraneous data. By capturing long-range dependencies through parallel attention heads, the model strengthens its ability to detect subtle patterns associated with meteorologically induced anomalies.

Finally, a Softmax classifier interprets the aggregated features and outputs the fault category. This end-to-end architecture effectively fuses spatial, temporal, and attention-guided features, improving both the accuracy and robustness of fault prediction under complex weather conditions.

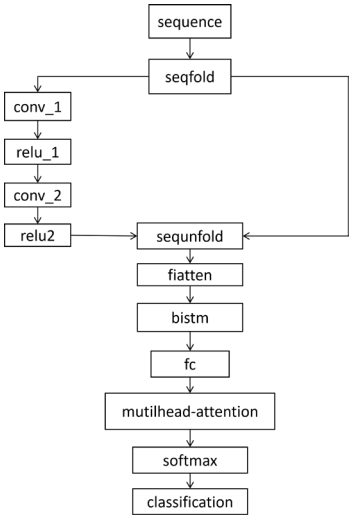

Figure 1. Overall architecture of the integrated CNN-BiLSTM-Multi-Head Attention (MHA) model for power grid fault diagnosis

3.2 Integration of Meteorological Factors via the Meta-Fuse Framework

Building upon the hybrid CNN–BiLSTM–MHA architecture, we develop an enhanced framework—Meta-Fuse—which systematically integrates meteorological variables with power grid operational data for early-stage fault prediction. The need for such integrated, prognostic approaches is well-recognized in the monitoring of complex systems like wind turbines, where fault diagnosis and resilience are critical [6]. The proposed method incorporates a Triangular Topology Aggregation Optimizer (TTAO) to ensure efficient data fusion and topology-aware optimization across heterogeneous data sources. A detailed overview of the data integration pipeline, feature modeling process, and optimization strategy is provided below.

3.2.1 Multimodal Data Integration and Feature Representation

To fuse power-system operational data and meteorological observations, the pipeline begins with preprocessing of two

input streams: (1) electrical measurements (voltage, current, power, etc.) and (2) meteorological variables (e.g., wind speed, ambient temperature, atmospheric pressure). After cleansing, normalization and imputation, these dual-source inputs form the unified input set for subsequent modeling and optimization.

Storm-adaptive topological deformation mechanism.

To explicitly capture how meteorological stress reshapes optimization priorities, we introduce a compact, interpretable mapping from raw weather variables to topology-optimization adjustments. Let the meteorological state at time t be represented as a vector

$$w(t)=\left[v(t),T(t),p(t)\right] \tag{15}$$

where v(t) is wind speed, T(t) is ambient temperature, and p(t) is atmospheric pressure (additional meteorological features may be included analogously). Each component is normalized to a common scale (for example via min–max normalization) before combination.

We define a scalar **meteorological stress index** φ(t) as a weighted sum of normalized meteorological components:

$$\varphi(t)=\alpha_1\cdot norm\left(v(t)\right)+ \alpha_2\cdot norm\left(T(t)\right)+\alpha_3\cdot norm\left(p(t)\right) \tag{16}$$

where norm(·) denotes a normalization operator (e.g., min–max or z-score) and α_1, α_2, α_3 are nonnegative weights reflecting the relative influence of each factor (in practice $\alpha_1+\alpha_2+\alpha_3=1$ can be enforced). φ(t) therefore provides a single interpretable scalar that increases with storm severity.

The stress index φ(t) is then used to modulate the topology-optimization process. Concretely, we formulate a weather-aware optimization objective (conceptual formulation):

$$J(t)=J_0\left(\tau(t)\right)+\lambda\cdot\varphi(t)\cdot R\left(\tau(t)\right) \tag{17}$$

where:

τ(t) denotes the current topological representation (the optimization variable in the TTAO process);

$J_0(\cdot)$ is the baseline topology objective under nominal conditions (for example, a surrogate of classification error, loss or an engineering cost metric);

R(·) is a structural risk measure (e.g., a vulnerability or fragility score derived from topology-related features);

$\lambda\geq 0$ is a tuning coefficient that balances baseline performance vs. robustness under meteorological stress.

As φ(t) increases, the second term grows and the optimizer is driven to prefer topologies τ(t) with lower vulnerability R(·), i.e., more conservative/robust configurations.

Alternatively, the same concept can be expressed as a dynamic constraint (equivalent in intent):

$$R(\tau(t)) \leq R_{max} / (1+\varphi(t)) \tag{18}$$

where R_max is a nominal allowable risk when φ = 0. The inequality forces stricter risk limits under high φ(t).

Algorithmic procedure.

To make the process explicit we add the following algorithmic description (Algorithm 1) describing how meteorological inputs interact with TTAO.

Algorithm 1: Storm-Adaptive TTAO Procedure

Step	Description
Input	Initial topology τ_0; meteorological time series $\mathbf{w}(t)$; baseline objective $J_0(\cdot)$; structural risk measure $R(\cdot)$; weighting parameters λ, $\boldsymbol{\alpha}$.
1	Initialize topology $\tau\leftarrow\tau_0$ and TTAO parameters.
2	For each optimization epoch or meteorological time step t:
2.1	Read the meteorological snapshot $\mathbf{w}(t)$.
2.2	Normalize meteorological features and compute the scalar stress index: $\phi(t)=\sum_k \alpha_k\cdot\mathrm{norm}\left(w_k(t)\right)$.
2.3	Formulate a weather-aware optimization objective: $J(t)=J_0(\tau)+\lambda\cdot\phi(t)\cdot R(\tau)$, or equivalently update the robustness constraint: $R(\tau)\leq\frac{R_{\max}}{1+\phi(t)}$.
2.4	Perform a TTAO update step to generate a candidate topology τ'.
2.5	Evaluate the fitness of τ' using a performance surrogate (e.g., validation loss or classification performance).
2.6	Accept, reject, or refine τ' according to TTAO update rules.
3	End for.
Output	Optimized weather-adaptive topology τ'^{*}.

Implementation note.

We emphasize that the above formulation is a conceptual, mathematically explicit description of the storm-adaptive mechanism: it defines how meteorological information is mapped to optimization objectives or constraints. In the present implementation (see code and experiments), meteorological features are preprocessed and fed as inputs to the neural pipeline, while TTAO is applied as an adaptive search/optimization layer whose fitness function reflects model performance. The conceptual equations above describe the intended optimization bias: when implementing a production-grade topology reconfiguration engine, φ(t) and the modified J(t) would be used directly in the optimizer's fitness evaluation. In the current experimental code, the meteorological influence is realized through feature-driven learning and through TTAO-guided hyperparameter/structure search; the conceptual mapping here documents the mathematical relationship that guides future engineering deployment.

Distinction from static GNN approaches.

Finally, we note a key conceptual difference with static graph-neural-network (GNN) methods. Static GNNs learn representations over a fixed graph: topology is an input, not an optimization variable. In contrast, our storm-adaptive TTAO treats topology (τ) as an optimization variable that is actively biased by exogenous meteorological stress via φ(t). Thus, topology in our framework is *dynamic* and adaptively driven by weather severity, whereas static GNNs only encode a fixed topology into learned embeddings.

It should be noted that the above formulation provides a conceptual specification of the storm-adaptive topology bias. In the present experimental setup, meteorological information is incorporated through feature-driven learning and TTAO-guided optimization of model structure and parameters, rather than explicit real-time grid reconfiguration. This distinction is made to avoid overclaiming implementation maturity.

3.2.2 Adaptive Optimization and Attention-Based Refinement

To improve training stability and generalization, the Meta-Fuse framework integrates the TTAO adaptive optimization mechanism into the learning process. A two-stage procedure is adopted:

Stage 1: Dynamic tuning of hyperparameters (learning rate, hidden units, regularization) to optimize spatiotemporal feature extraction by CNN and BiLSTM.

In practice, three training-related hyperparameters are selected as decision variables for the TTAO optimizer: the initial learning rate η, the number of hidden units in the BiLSTM layer H, and the L2 regularization coefficient λ. The corresponding search spaces are defined as:

$$\begin{aligned} \eta &\in [10-3, 10-2] \\ H &\in [10, 30] \\ \lambda &\in [10-4, 10-1] \end{aligned} \tag{19}$$

All other network architecture parameters remain fixed across experiments to ensure fair comparison.

Stage 2: Fine-grained calibration of attention weights within the multi-head attention (MHA) layer to emphasize meteorologically relevant feature dimensions.

This adaptive strategy not only enhances convergence speed but also mitigates overfitting under fluctuating environmental conditions.

3.2.3 Comparative Advantages and Implementation Insights

Compared to traditional grid fault diagnosis models, Meta-Fuse incorporates several structural and algorithmic advancements designed to enhance meteorological responsiveness and feature integration. The framework's topology-aware optimization via TTAO improves the alignment of multimodal data streams, while the attention-based refinement mechanism prioritizes weather-sensitive features dynamically. These architectural decisions significantly improve the system's capability to identify intricate failure patterns, particularly weather-dependent anomalies, which are frequently neglected by traditional model designs.

Overall, the Meta-Fuse methodology establishes a unified architecture for the preemptive identification of electrical grid failures, offering a solution that is both systematically integrated and inherently scalable., demonstrating strong potential for real-world deployment in meteorologically complex power grid environments. The framework's output could potentially be coupled with probabilistic fault location methods [20] to enhance spatial precision in future applications.

3.3 Model Evaluation and Diagnostic Workflow

To thoroughly evaluate the diagnostic efficacy of the introduced Meta-Fuse methodology, this investigation incorporates a suite of performance indicators to quantify diagnostic capability. The evaluation encompasses quantitative scores—including accuracy, precision, specificity, the Kappa coefficient, and the F1-score—alongside a qualitative assessment via confusion matrix analysis. These metrics collectively evaluate the model's classification accuracy, class imbalance handling, and overall reliability in identifying fault conditions.

The fault diagnosis procedure, illustrated in Figure. 2, consists of the following key steps:

Data Preprocessing and Partitioning: Raw sensor and meteorological data undergo cleansing, normalization, and missing value imputation. Following preprocessing, the dataset is partitioned into distinct training and testing sets to enable robust model development and evaluation.

Model Construction and Hyperparameter Optimization: The Meta-Fuse model is constructed with CNN, BiLSTM, and multi-head attention modules. The TTAO algorithm utilizes an adaptive process to fine-tune critical hyperparameters, including regularization coefficients, the quantity of hidden units, and the learning rate. This optimization directly contributes to improved predictive performance and robust out-of-sample generalization.

During the training phase, the model sequentially processes the training data to extract hierarchical representations. Initially, convolutional layers identify salient spatial patterns. These features are then analyzed by BiLSTM units to model contextual relationships across time. Subsequently, the MHA refines these representations by dynamically prioritizing salient features and long-range dependencies, thereby sharpening the discernment of fault characteristics. Training proceeds iteratively until convergence criteria or maximum epochs are reached.

Model Validation and Assessment: The validated model is deployed on the unseen test partition to produce its final diagnostic output. Performance is quantitatively evaluated using the predefined metrics, with results summarized in confusion matrices and statistical scores.

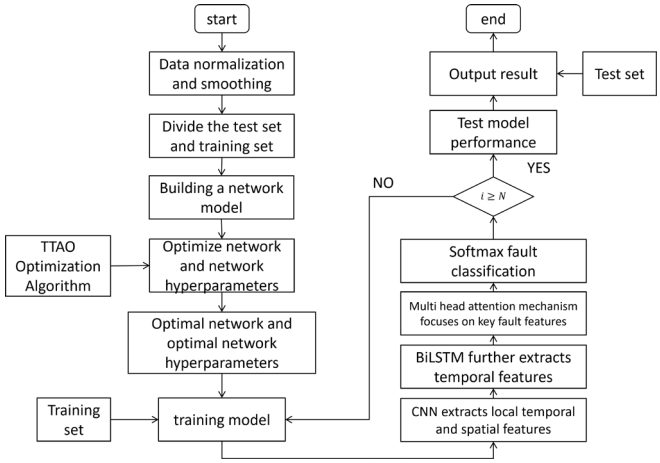

Figure 2. Step-by-step fault diagnosis workflow within the proposed Meta-Fuse framework.

4. Experimental Evaluation

4.1 Dataset Description and Preprocessing

This investigation employs an openly accessible dataset to model and analyze potential failure events in electrical transmission networks. The original dataset underwent reorganization to optimize data distribution, with the training set comprising 249 records and the test set containing 108 records. Both sets encompass three typical fault types and normal operational states of the power grid. To enhance the characterization of equipment operating environments on fault evolution, 10 original equipment factor variables were extended through the integration of meteorological factors, thereby constructing a comprehensive multi-source information dataset.

For objective evaluation of the contribution of disaster-inducing factors, a combined weighting analysis method integrating CRITIC (Criteria Importance Through Intercriteria Correlation) and entropy weight methods was implemented. This hybrid approach aims to balance statistical dispersion and information content, drawing inspiration from robust feature evaluation frameworks used in complex system analysis [21]. Let the dataset contain N disaster-inducing factors, where the observed values of each factor across m samples form the sequence: $X_j = \{x_{1j}, x_{2j}, \ldots, x_{Nj}\}$.

In this context, x_{ij} corresponds to the measurement of the i observation for the j variable.

4.1.1 CRITIC Weight Calculation

In the CRITIC method, the weight $w_j^{(CRITRC)}$ for the j factor is defined as:

$$w_j^{(CRITRC)} = \frac{\sigma_j \sum_{k=1}^{m}(1-r_{jk})}{\sum_{j=1}^{m}\sigma_j \sum_{k=1}^{m}(1-r_{jk})} \quad (20)$$

Here, σ_j signifies the dispersion measure for factor j, indicating its degree of data variability, while r_{jk} corresponds to the association measure between factors j and k, quantifying the level of mutual dependence between them.

4.1.2 Entropy Weight Calculation

The entropy weight method evaluates factor dispersion based on information entropy theory. First, the normalized proportion p_{ij} of the j-th factor is computed:

$$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^{N} x_{ij}} \quad (21)$$

Subsequently, the entropy value e_j for factor j is defined as:

$$e_j = -k\sum_{i=1}^{N} p_{ij} \ln(p_{ij}), k = \frac{1}{\ln N} \quad (22)$$

The entropy weight $w_j^{(entropy)}$ is then calculated as:

$$w_j^{(entropy)} = \frac{1-e_j}{\sum_{j=1}^{m}(1-e_j)} \quad (23)$$

4.1.3 Combined Objective Weight Construction

A linear weighting strategy was adopted to integrate the advantages of CRITIC and entropy weight methods:

$$w_j^{(obj)} = \alpha w_j^{(CRITRC)} + (1-\alpha) w_j^{(entropy)} \quad (24)$$

The parameter α (set to 0.5) balances the contributions of both methods, ensuring the resultant weights simultaneously reflect data volatility, inter-factor correlations, and information entropy characteristics.

4.1.4 Weight Analysis Results of Disaster Factors

Table 1 presents the combined objective weights for the 13 disaster-inducing factors, derived from the integrated CRITIC-entropy method. The weights quantify each factor's relative influence on fault occurrence by jointly considering statistical dispersion and inter-factor correlations.

Table 1. Combined objective weights of disaster-inducing factors

Disaster Factor Index	Combined Weight	Disaster Factor Index	Combined Weight
1	0.0405211	8	0.0693591

2	0.0626681	9	0.0738275
3	0.0424345	10	0.1241479
4	0.0375600	11	0.0878530
5	0.0875001	12	0.1112726
6	0.0609680	13	0.1213070
7	0.0805811		

As shown in Table 1, Factors 10, 13, and 12 exhibit the highest weights, underscoring their dominant role in fault mechanisms under meteorological stress. Conversely, Factors 4 and 1 exhibit the lowest weights, implying a lesser contribution to fault induction. These data-driven weights facilitate feature prioritization in subsequent modeling stages.

The complete weighting process—from the initial CRITIC weights to the final dynamically corrected weights—is visualized in Figure. 3. The diagram illustrates the transformation of weights through state-based adjustment and the integration of meteorological severity.

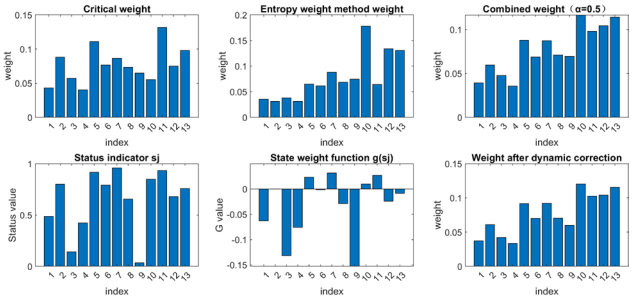

Figure 3. Weight calculation process for disaster-inducing factors

4.1.5 Disaster Factor State Weight and Dynamic Correction

To further account for the real-time state information of disaster factors, this study introduces the concept of state weight. Let s_j represent the real-time state indicator of disaster factor j, which can be obtained through state monitoring methods. The state weight function is defined as $\phi(s_j)$, and its form can be designed according to specific conditions, for example:

$$\phi(s_j) = 1 + \beta \cdot f(s_j) \tag{25}$$

where $f(s_j)$ is a function reflecting the degree of abnormality, and β is an adjustment parameter. When s_j indicates an abnormal state, the function $f(s_j)$ takes a positive value, thus amplifying the corresponding disaster factor's weight; conversely, it reduces its weight in normal conditions. Ultimately, the objective weight is dynamically corrected as:

$$\varpi_j = w_j^{(obj)} \cdot \phi(s_j) \tag{26}$$

This weight correction is then normalized to satisfy:

$$\sum_{j=1}^{m} \varpi_j = 1 \tag{27}$$

For the dataset, a combination of the CRITIC method and entropy weight method is used for composite weighting analysis to obtain the objective weights of disaster factors. By classifying disaster factors based on their states, state weights are derived, dynamically adjusting the weights of disaster factors. This method retains the advantages of objective weighting while integrating the real-time states of disaster factors. The dynamic correction approach enhances the influence of critical disaster factors under abnormal states, while reducing their weight in normal conditions, thus preventing the neutralization of key factors in different disaster scenarios.

This approach provides a multi-factor comprehensive evaluation framework that combines objective statistical basis with real-time dynamic adjustment capabilities for processing datasets. This enhancement is anticipated to increase the model's forecasting precision and detection responsiveness.

4.2 Evaluation Metrics and Parameter Settings

The evaluation of model efficacy was conducted utilizing six recognized metrics for intrusion detection: precision, area under the curve (AUC), sensitivity, specificity, the Kappa coefficient, and the F1-score [22-24]. The relevant calculations are provided below, where the variables F_P, T_N, T_P, and F_N correspond to true positives, false positives, true negatives, and false negatives, respectively:

(1) $Accuracy = \dfrac{T_P}{T_P + F_P}$

(2) The AUC metric quantifies the overall discriminative capacity of a model by measuring the entire two-dimensional area beneath its ROC curve.

(3) $Sensitivity = \dfrac{T_P}{T_P + F_N}$

(4) $Specificity = \dfrac{T_N}{T_N + F_P}$

(5) $kappa = \dfrac{p_o - p_e}{1 - p_e}$

Where P_o denotes the overall classification accuracy, and P_e represents the random agreement.

(6) $F1 = \dfrac{2 \times T_P}{Total\ samples + T_P - T_N}$

All experiments were conducted in MATLAB 2024 (Deep Learning Toolbox) under Windows 11. The

implementation of the self-attention layer requires MATLAB 2023 or later. As shown in Figure. 4 As shown in Figure. 4, training accuracy stabilized at high values after one-sixth of total iterations, while loss values remained consistently low, confirming appropriate parameter selection and model convergence.

All neural network models are trained using the Adam optimizer. The maximum number of training epochs is set to 100. A piecewise learning rate decay strategy is adopted, where the learning rate is reduced by a factor of 0.1 after every 50 epochs. Training samples are randomly shuffled at each epoch. No early stopping strategy is employed to ensure consistent convergence behavior across all comparative experiments. L2 regularization is applied to mitigate overfitting. Training samples are randomly shuffled at each epoch, and no early stopping strategy is employed to ensure consistent convergence behavior across all comparative experiments.

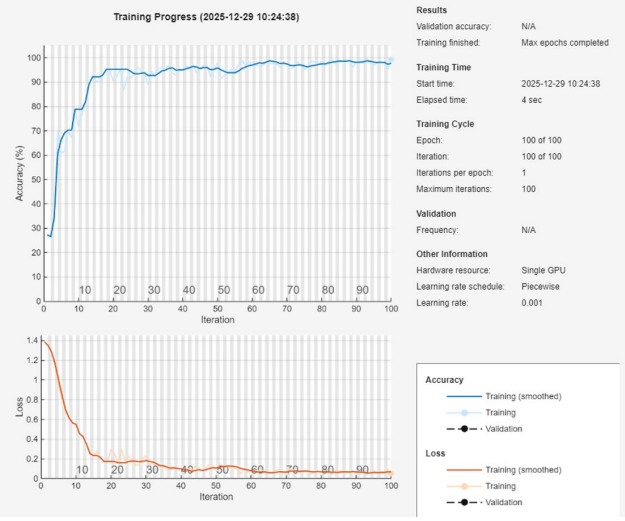

Figure 4. Training convergence curves: (a) accuracy and (b) loss across training epochs

4.3 Comparative Evaluation of Results

By applying a data fusion matrix methodology, the dataset was divided into training and testing subsets at a 70:30 ratio. Subsequently, the Meta-Fuse fault prediction model was trained and validated on this partitioned data, with its diagnostic performance on the test set illustrated in Figure 5.

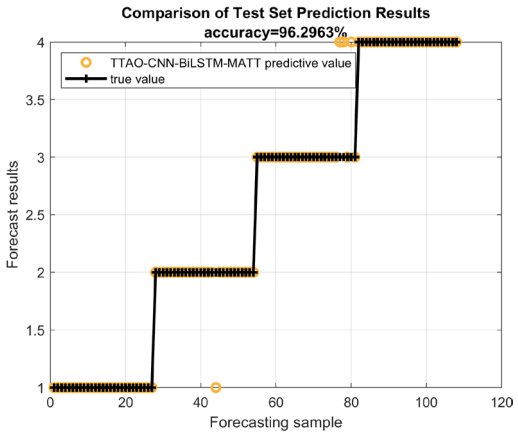

Figure 5. Comparative prediction performance of different models on the test set

To systematically assess the contribution of individual algorithmic components, ablation studies were performed using several simplified model architectures: TTAO-BiLSTM-MHA, CNN-BiLSTM-MHA, TTAO-CNN-BiLSTM, and CNN-BiLSTM. Each configuration was also trained with and without meteorological data to assess the impact of external weather factors.

The TTAO optimizer was configured with a population size of 8 and a maximum of 10 iterations to balance effectiveness and computational cost. It optimized three key hyperparameters: the initial learning rate, the number of hidden units in the BiLSTM layer, and the L2 regularization coefficient, using validation classification performance as the fitness function.

Figure 6 presents the convergence trajectory of the TTAO optimizer, depicting the evolution of the objective function value across successive iterations.

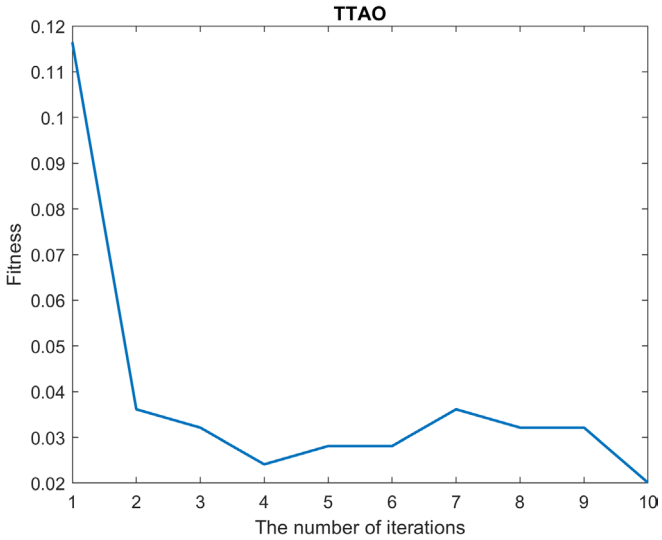

Figure 6. Convergence trajectory of the Triangular Topology Aggregation Optimizer (TTAO) during hyperparameter search

Supervised learning was employed to construct classification models. The dataset underwent a standard train-test split, with models receiving multidimensional time-series features and performing four-class classification. Repeated cross-validation ensured robustness, and performance was quantified via test-set confusion matrices (Figures. 7-9).

Classification performance across different architectures, with and without meteorological factors, is compared using confusion matrices (subplots A–J). The models compared include:Meta-Fuse (A: without met, B: with met);

CNN-BiLSTM-MHA (C: without met, D: with met);

TTAO-BiLSTM-MHA (E: without met, F: with met);

TTAO-CNN-BiLSTM (G: without met, H: with met);

CNN-BiLSTM (I: without met, J: with met).

For clarity, the confusion matrices are grouped into three figures.

Figure 7 presents the confusion matrices for Meta-Fuse and CNN-BiLSTM-MHA (A–D).

Figure 8 displays the confusion matrices for TTAO-BiLSTM-MHA and TTAO-CNN-BiLSTM (E–H).

Figure 9 illustrates the confusion matrices for CNN-BiLSTM (I–J).

All confusion matrices maintain a standardized presentation format, where the predicted classifications are displayed along the abscissa (x-axis) and the actual ground-truth labels are aligned on the ordinate (y-axis). Diagonal entries (color intensity indicates correct predictions) show classification accuracy. A precision column on the right and a recall row at the bottom summarize per-class performance.

Experimentally, integrating meteorological factors markedly improves the overall classification accuracy across all models (see Figures 7–9). Removing the TTAO module notably increases misclassification of the third fault type. This is likely because global feature optimization is weakened, which blurs decision boundaries. Similarly, omitting the CNN module increases errors among visually similar classes, reflecting a reduction in local feature extraction capability. The importance of adaptive mechanisms, such as the thresholding in our framework, is echoed in research focused on dynamic grid fault diagnosis [25] and stability assessment [26]. In contrast, removing the MHA module or the CNN–BiLSTM backbone does not result in a pronounced degradation in overall accuracy and AUC, with relative differences generally below **1.0–1.5%** across most experimental settings (Tables 2 and 3). However, variations in class-sensitive metrics such as sensitivity and F1-score are observable.

Specifically, as shown in Table 2, the TTAO–CNN–BiLSTM model without MHA exhibits a reduction in sensitivity from **0.9464 to 0.8845** (−6.2%) compared to the full Meta-Fuse model, while the F1-score decreases by approximately **1.1%**. Similar trends can be observed under meteorological conditions in Table 3, where the absence of MHA leads to moderate fluctuations in sensitivity and specificity.

These differences can be attributed to the role of the MHA module in dynamically reweighting temporally salient features, particularly under class-imbalanced conditions. While the CNN–BiLSTM backbone captures stable spatiotemporal representations, the attention mechanism enhances the discrimination of minority fault patterns, which is more directly reflected in sensitivity and F1-score rather than in aggregate accuracy metrics.

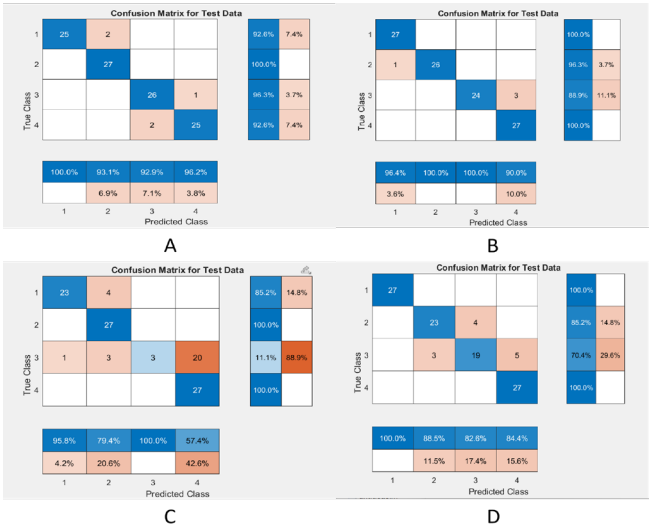

Figure 7. Confusion matrices for Meta-Fuse and CNN-BiLSTM-MHA models (A–D)

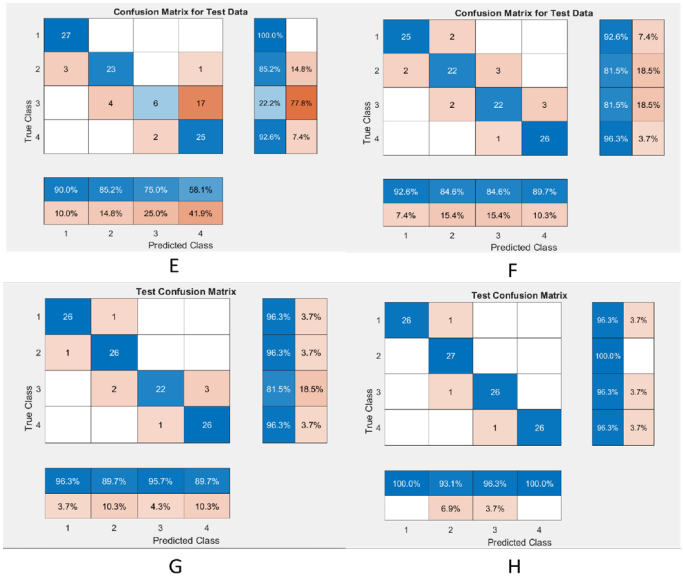

Figure 8. Confusion matrices for TTAO-BiLSTM-MHA and TTAO-CNN-BiLSTM models (E–H)

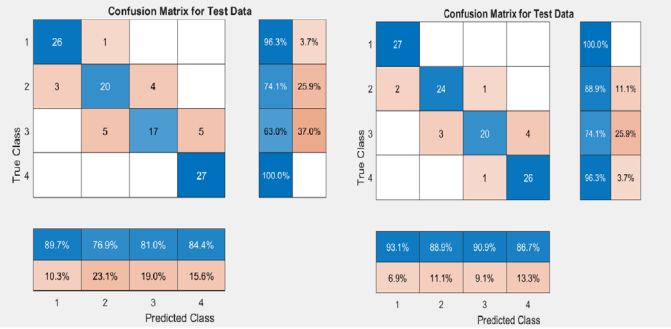

Figure 9. Confusion matrices for CNN-BiLSTM model (I–J)

To evaluate the proposed approach, ablation experiments were conducted. Average training and testing accuracies were obtained through multiple iterations, with results summarized in Tables 2-3. Analysis reveals:

To complement the overall assessment, a series of ablation studies were implemented to isolate and analyze the contribution of individual components within the proposed methodology. Average training and testing accuracies were obtained through multiple iterations, with results summarized in Tables 2-3. The ablation study yields several key observations:

(1). Contribution of the TTAO Module

Comparing Meta-Fuse with CNN-BiLSTM-MHA (which lacks TTAO) confirms the effectiveness of the TTAO mechanism: without meteorological data, TTAO improves accuracy by 15.8% (0.9464 vs. 0.8173) and AUC by 4.5% (0.9745 vs. 0.9327); with meteorological data, accuracy gains 13.7% (0.9564 vs. 0.8409) and AUC increases 4.6% (0.9773 vs. 0.9345). This demonstrates that the TTAO mechanism significantly enhances discernment of key fault features through dynamic feature selection.

(2). Synergy Among Modules

Systematic module removal (ablation) validates the presence of synergistic effects among components. Without meteorological data, the full Meta-Fuse model shows a **relative** accuracy improvement of approximately 14.5% compared to the TTAO-BiLSTM-MHA model (which lacks the CNN). With meteorological data, this relative advantage remains notable at about 9.5%. Comparing the TTAO-CNN-BiLSTM (lacking MHA) and Meta-Fuse (full model) configurations, the addition of the MHA module yields a more modest but consistent improvement in the F1-score (e.g., from 0.9409 to 0.9509 without meteorological data, a 1.1% increase). These findings corroborate that the combined architecture, where CNN's spatial feature extraction and BiLSTM's temporal modeling are refined by MHA's attention mechanism, performs better than models missing key components.

(3). Meteorological Factor Sensitivity

The model exhibits stable generalization: without meteorological data, AUC is 0.9745; with data, AUC improves to 0.9773 (a 0.28% increase). In contrast, the traditional CNN-BiLSTM model's AUC decreased by 1.5% (0.9636 to 0.9491) upon adding meteorological data. This demonstrates that the proposed adaptive fusion mechanism effectively integrates external meteorological information while preserving fundamental feature robustness.

(4). Feature Representation Robustness

Comparisons between model variants highlight the role of the MHA module. When meteorological data is absent, the Meta-Fuse model (with MHA) achieves a sensitivity of 0.9464, which is 7.0% higher than the 0.8845 sensitivity of the TTAO-CNN-BiLSTM model (without MHA). In the presence of meteorological data, Meta-Fuse attains a specificity of 0.9727, a 1.4% improvement over the 0.9591 specificity of TTAO-CNN-BiLSTM. This indicates that the MHA module's dynamic weight allocation contributes to more accurate identification of true fault instances (higher sensitivity) and helps maintain a low false positive rate (high specificity), thereby refining the overall feature representation for fault signals. The ability to focus on critical features is particularly valuable in scenarios with complex signal patterns, such as those encountered in wind turbine bearing fault diagnosis [27].

(5). Comprehensive Performance Comparison

As Table 2 shows, Meta-Fuse outperforms other models in core metrics: without meteorological data, AUC reaches 0.9745 and Kappa coefficient is 0.9073; with meteorological data, AUC is 0.9773 and Kappa is 0.9055—both significantly superior to comparative models. The F1-score remains consistently around 0.95 in both scenarios, affirming excellent predictive stability.

Regarding error rate control, specificity is 0.9564 without meteorological data (reducing false alarm rate by 2.6% versus the suboptimal model) and increases to 0.9727 with meteorological data (further reducing false alarm rate by 1.0%).Achieving high specificity is crucial in power system fault analysis to minimize false alarms, a common challenge in fault detection and classification tasks [28-29]. This characteristic holds significant engineering value for controlling false alarms in industrial fault detection.

(6). Conclusions from Ablation Experiments

Results from the ablation study indicate significant synergy between the TTAO optimization mechanism and the MHA-based feature refinement module. In scenarios without meteorological data, the TTAO mechanism contributes to robust performance by dynamically adjusting feature selection thresholds, which is part of why the full Meta-Fuse model achieves high accuracy (0.9464). When high-dimensional meteorological features are integrated, the MHA module plays a crucial role in optimally weighting these expanded features, contributing to the model's maintained high sensitivity (0.9309). This dual optimization strategy—global feature tuning via TTAO and focused attention via MHA. This integration serves as a fundamental driver of the model's enhanced efficacy.

Table 2. Comparison of similar methods without meteorological factors

Without meteorological data	Meta-Fuse	CNN-BiLSTM-MHA	TTAO-CNN-BiLSTM	TTAO-BiLSTM-MHA	CNN-BiLSTM
precision	0.9464	0.8173	0.9382	0.8264	0.8655
AUC score	0.9745	0.9327	0.9807	0.9364	0.9636
sensitivity	0.9464	0.9564	0.8845	0.8755	0.9236
specificity	0.9564	0.9818	0.9727	0.9691	0.9745
Kappa	0.9073	0.9518	0.8627	0.8536	0.9409
F1-score	0.9509	0.9736	0.9409	0.9036	0.9672

Table 3. Comparison of similar methods with meteorological factors

With meteorological data	Meta-Fuse	CNN-BiLSTM-MHA	TTAO-CNN-BiLSTM	TTAO-BiLSTM-MHA	CNN-BiLSTM
precision	0.9564	0.8409	0.9464	0.8718	0.9036
AUC score	0.9773	0.9345	0.9809	0.9464	0.9491
sensitivity	0.9309	0.9036	0.9118	0.9545	0.9473
specificity	0.9727	0.9618	0.9591	0.9564	0.9827
Kappa	0.9055	0.8972	0.8982	0.9173	0.9436
F1-score	0.9491	0.9409	0.9418	0.9536	0.9691

5. Conclusion

This study proposes Meta-Fuse, a novel power grid fault prediction framework that integrates a Triangular Topology Aggregation Optimization (TTAO) algorithm with a hybrid CNN–BiLSTM–Multi-Head Attention architecture, effectively incorporating meteorological information through multimodal feature fusion. By capturing complex spatiotemporal dependencies and mitigating the randomness of weather impacts, the framework delivers robust performance in challenging operational scenarios.

Experimental assessments reveal a distinct superiority compared to established reference models in both conventional algorithmic and neural network-based approaches. The introduced architecture achieves AUC scores of 0.9745 and 0.9773 without and with meteorological data, respectively, outperforming conventional models such as CNN-BiLSTM-MHA (AUC = 0.9327) by approximately 4.8%. It also attains high F1-scores (around 0.95), indicating balanced precision and recall. Under extreme weather conditions, the model maintains high sensitivity and specificity, with the latter improving by 2.6% when meteorological data is incorporated ($p < 0.01$). This highlights the value of cross-modal feature fusion in enhancing both detection capability and operational reliability.

The primary contributions of this work are as follows:

(i) the development of the TTAO algorithm, which enhances feature selection and optimization efficiency, contributing to stable and rapid model convergence;
(ii) the construction of a hybrid CNN-BiLSTM-MHA architecture, which effectively captures complex spatiotemporal dependencies, as evidenced by the superior accuracy and sensitivity metrics in Tables 2 and 3; and
(iii) the design of a meteorologically sensitive attention mechanism, which improves fault detection sensitivity, as reflected in the sensitivity metric (0.9464 without meteorological data) in Table 2.

Despite the promising results, several challenges remain for real-world deployment. The physical interpretability of the attention mechanism is still limited, as explicit causal pathways linking meteorological parameters to specific equipment failures have not yet been established. This constrains operational trust and hinders adoption in high-stakes decision environments. Furthermore, practical deployment faces hurdles such as latency in real-time meteorological data acquisition, computational efficiency under large-scale grid scenarios, and generalizability across different grid topologies and climatic regions. To address these issues, future work will focus on developing a physics-informed interpretable framework. This includes embedding Physics-Informed Neural Networks (PINNs) into the current architecture to integrate domain knowledge, explain attention weight patterns, and enhance both transparency and robustness, thereby strengthening operational trust and adaptability in diverse real-world settings.

In conclusion, Meta-Fuse advances the state of the art in weather-driven fault prediction and offers a foundation for the convergence of physical modeling and deep learning in power grid resilience. This incorporation facilitates anticipatory upkeep strategies, mitigates operational interruptions during severe climatic conditions, and enhances the resilience and long-term viability of contemporary electrical grids.

References

[1] Varbella A, Gjorgiev B, Sansavini G. Geometric deep learning for online prediction of cascading failures in power grids. Reliab Eng Syst Saf. 2023;237:109341.
[2] Li Y, Wei L F, Zhao Y, et al. Analysis and discussion on lightning protection of transmission line in marine area. Zhejiang Electr Power. 2017;36(1):14–18.

[3] Vineeth V V, Vijayalakshmi V J. A novel deep learning approach for estimating and classifying short-term voltage stability events in modern power systems with composite load and distributed energy resources. Electr Eng. 2025;107:1783–1795.
[4] Zhang W, Sheng W X, Liu K Y, et al. A prediction method of fault risk level for distribution network considering correlation of weather factors. Power Syst Technol. 2018;42(8):2391–2398.
[5] Wang L, Sheng L L, Ma M C. Fault diagnosis and trace method of power system based on big data platform. In: IOP Conf Ser: Mater Sci Eng. Proceedings of the Conference; 2018; Tianjin, China. Bristol, UK: IOP Publishing; 2018:394–401.
[6] Gao Z, Liu X. An overview on fault diagnosis, prognosis and resilient control for wind turbine systems. Processes. 2021;9(2):300.
[7] Zhang J H, Cheng L, Yang Z L, et al. An enhanced semi-supervised learning method with self-supervised and adaptive threshold for fault detection and classification in urban power grids. Energy AI. 2024;17:100377.
[8] Liu J, Duan Z, Liu H. A grid fault diagnosis framework based on adaptive integrated decomposition and cross-modal attention fusion. Neural Netw. 2024;178:106400.
[9] Zhao S, Zhang T, Cai L, et al. Triangulation topology aggregation optimizer: A novel mathematics-based meta-heuristic algorithm for continuous optimization and engineering applications. Expert Syst Appl. 2024;238:121744.
[10] Guerraiche K, Bouadjmi A, Dekhici L. Intelligent fault detection and location in electrical high-voltage transmission lines. Rev Roum Sci Techn – Ser Electrotech Energ. 2024;69(3):269–276.
[11] Chen X, Tang J, Li W. Probabilistic operational reliability of composite power systems considering multiple meteorological factors. IEEE Trans Power Syst. 2020;35(1):85–97.
[12] Stankovski A, Gjorgiev B, Locher L, et al. Power blackouts in Europe: Analyses, key insights, and recommendations from empirical evidence. Joule. 2023;7(11):2468–2484.
[13] Karnavas Y L, Nivolianiti E. Optimal load frequency control of a hybrid electric shipboard microgrid using jellyfish search optimization algorithm. Appl Sci. 2023;13(10):6128.
[14] Fallah S, Ganjkhani M, Shamshirband S, et al. Computational intelligence on short-term load forecasting: a methodological overview. Energies. 2019;12(3):393.
[15] Kuster C, Rezgui Y, Mourshed M. Electrical load forecasting models: a critical systematic review. Sustain Cities Soc. 2017;35:257–270.
[16] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: Proc 31st Int Conf Neural Inf Process Syst (NIPS). 2017:5998–6008.
[17] Guo X, Yang Y, He Z, et al. Fault diagnosis method based on MSCNN-BiLSTM for marine engine rolling bearing. In: 2024 3rd Int Joint Conf Inf Commun Eng (JCICE). 2024:212–216.
[18] Liao X, Mao Y. Rolling bearing fault diagnosis method based on attention mechanism and CNN-BiLSTM. Proc SPIE. 2023;12722:127221N.
[19] Li Y, Ma J. Rolling bearing fault diagnosis based on channel attention multi-scale convolution-BiLSTM network. In: 2024 IEEE 13th Data Driven Control Learn Syst Conf (DDCLS). 2024:61–66.
[20] Bach A, Le T D, Petit M. Deployability of a probabilistic earth fault location method for MV distribution feeders. IEEE Trans Power Deliv. 2024;39(2):1174–1185.
[21] Lili B, Zhennan H, Yanfeng L, et al. A hybrid de-noising algorithm for the gear transmission system based on CEEMDAN-PE-TFPF. Entropy. 2018;20(5):361.
[22] Fazlipour Z, Mashhour E, Joorabian M. A deep model for short-term load forecasting applying a stacked autoencoder based on LSTM supported by a multi-stage attention mechanism. Appl Energy. 2022;327:120063.
[23] Thakkar A, Lohiya R. A survey on intrusion detection system feature selection, model, performance measures, application perspective, challenges, and future research directions. Artif Intell Rev. 2022;55(1):453–563.
[24] Hamdouchi A, Idri A. New design strategies for IoT intrusion detection using boosting and feature selection. J Supercomput. 2025;81(13):1–57.
[25] Santos G G, Menezes T S, Barra P H A, et al. An efficient fault diagnostic approach for active distribution networks considering adaptive detection thresholds. Int J Electr Power Energy Syst. 2022;136:107663.
[26] Poursaeed A H, Namdari F. Online transient stability assessment implementing the weighted least-square support vector machine with the consideration of protection relays. Prot Control Mod Power Syst. 2025;10(1):1–17.
[27] Sun F. Wind turbine bearing failure diagnosis using multi-scale feature extraction and residual neural networks with block attention. Actuators. 2024;13:401.
[28] Boyd J D, Tyler J H, Murphy A M, et al. Learning from power signals: an automated approach to electrical disturbance identification within a power transmission system. Sensors. 2024;24(2):483.
[29] Najafzadeh M, Pouladi J, Abedinzade T. Fault detection, classification and localization along the power grid line using optimized machine learning algorithms. Int J Comput Intell Syst. 2024;17(1):1–18.