

Spatiotemporal Feature Extraction from Integrated Multivariate Time Series via MCNN-LSTM for Wind Turbine Fault Diagnosis

Xiangjun Zeng, Xiangqing Fang, Binqiao Zhang, Chen Feng, and Shengyuan Zhou

Abstract—Wind turbine (WT) fault diagnosis using supervisory control and data acquisition (SCADA) data is challenged by severe class imbalance, which deteriorates the recognition of minority class. To address this issue, a spatiotemporal feature extraction framework from integrated multivariate time series (IMTS) is proposed, with coordinated designs in data representation, model architecture, and loss regularization. First, a dual-label IMTS is constructed to fuse multi-source SCADA data via sliding windows. Second, a spatiotemporal feature learning architecture is developed by integrating a multi-scale convolutional neural network (MCNN) and a stacked long short-term memory (LSTM) network and designing the MCNN-LSTM model. Finally, an improved weighted cross-entropy loss is proposed, which incorporates kernel density estimation-based sample probabilities as regularization, thereby directing model attention toward minority and hard-to-classify samples. Experimental results on a real SCADA dataset demonstrate that the proposed MCNN-LSTM model achieves a Macro-recall score of 0.943 and a G-mean score of 0.942, outperforming comparison models in comprehensive performance.

Index Terms—Wind turbine (WT), fault diagnosis, integrated multivariable time series (IMTS), spatiotemporal feature, cross-entropy loss, multi-scale convolutional neural network (MCNN), long short-term memory (LSTM).

I. INTRODUCTION

WORLDWIDE installation of wind turbines (WTs) has increased significantly over the past decades due to the clean and sustainable advantages of wind energy over traditional sources. As most WTs are installed in remote island and coastal areas, they are subjected to harsh operation environments and limited maintenance [1]. Consequently, WTs

are more prone to faults after a period of operation compared with traditional hydropower and thermal power generators. Unexpected faults may disrupt normal operation, even lead to prolonged downtime, seriously affecting power generation and incurring operation and maintenance costs. Therefore, accurately identifying and maintaining WT faults in a timely manner are of great significance for reducing operation and maintenance costs and improving economic benefits [2].

In recent years, data-driven fault diagnosis methods have attracted increasing research attention. These methods attempt to identify faults by mining fault-related information from WT condition monitoring data [3]. For example, [4] presented a conditional variational generative adversarial network (GAN) for WT bearing fault diagnosis. Reference [5] proposed an adaptive deep learning method for WT gearbox fault diagnosis. Reference [6] proposed a WT bearing fault diagnosis method based on variational mode decomposition of parameter optimization and multi-feature fusion. Reference [7] proposed a skewness-driven deep learning method for diagnosing DC-link bus faults in WT using rotor and stator currents. These methods have demonstrated excellent performance in fault diagnosis of critical WT components such as the gearbox, blade, and generator, relying on vibration signals, acoustic emission signals, and high-frequency current signals. However, not all WTs can provide these signals, as installing specialized sensors such as accelerometers and acoustic emission sensors requires additional investment.

Alternatively, most current wind power manufacturers have employed supervisory control and data acquisition (SCADA) system to continuously monitor the operation status and performance of WTs. A typical SCADA system usually records a substantial amount of sensory data, typically with sampling intervals ranging from 1 s to 10 min. These data collected at fixed intervals essentially form continuous time series, making temporal dependency a fundamental attribute of SCADA data. Over time, this accumulates a significant volume of operation data of WTs. These data contain meaningful monitoring and diagnostic information related to the health status of the WTs [8]. Consequently, data-driven fault diagnosis methods utilizing SCADA data have attracted increasing attention in recent years since no additional sensors or data acquisition devices are required. Among these, fault diagnosis methods based on deep learning algorithms

Manuscript received: May 13, 2025; revised: August 4, 2025; accepted: December 31, 2025. Date of CrossCheck: December 31, 2025. Date of online publication: February 27, 2026.

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

X. Zeng (corresponding author), X. Feng, and B. Zhang are with the College of Electrical Engineering & New energy, China Three Gorges University, Yichang, China, and they are also with Hubei Provincial Key Laboratory for Operation and Control of Cascaded Hydropower Station, China Three Gorges University, Yichang, China (e-mail: sxdxzxj@ctgu.edu.cn; 15908075013@163.com; zb@ctgu.edu.cn).

C. Feng is with the College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao, China (e-mail: fchen@sdust.edu.cn).

S. Zhou is with Jinan Power Supply Company of State Grid Shandong Electric Power, Jinan, China (e-mail: 453932802@qq.com).

DOI: 10.35833/MPCE.2025.000413



have garnered significant interest due to their ability to perform nonlinear representation and automatic extraction of fault-related features from massive data. For example, [9] and [10] combined deep learning with attention mechanisms to extract fault features from SCADA data for WT fault diagnosis. Reference [11] employed a long short-term memory (LSTM) network to model temporal dependencies in SCADA data and proposed a WT operation status calibration framework that integrated deep learning with fuzzy comprehensive evaluation to detect potential gearbox faults.

Beyond the intrinsic temporal dependencies in SCADA data, the cross-variable spatial correlations are equally critical for diagnosis. These interdependencies arise from the interconnected physics of WT subsystems, leading to observable correlations such as those between wind speed and power output, or between shaft speed and bearing temperature. Such systemic relationships across variables are conventionally termed as spatial features [12], [13], as they reflect how faults propagate through the spatially distributed assets. Therefore, numerous diagnostic methods have focused on constructing models capable of learning these spatial features for accurate fault diagnosis and localization. Reference [14] considered the potential spatial relationships between variables and proposed a WT fault diagnosis method based on an improved convolutional neural network (CNN). References [15] and [16] combined CNN and LSTM to improve the accuracy of WT fault diagnosis by learning spatial and temporal fault features from the raw SCADA data.

Although the deep learning-based WT fault diagnosis methods have achieved good performance, most of them assume that the number of different samples is roughly the same. However, it is crucial to acknowledge that WT faults occur as sporadic events, with varying probabilities among different fault types, and failed WTs cannot operate for long periods. As a result, the SCADA data contain significantly fewer fault samples than normal samples, and the quantities of different fault samples also differ substantially [17]. Training data imbalance can lead to bias in fault diagnosis models since knowledge derived from abundant data tends to dominate the training process, particularly favoring the majority class. Therefore, data imbalance is a widespread problem in practice and can negatively affect the fault diagnosis of WTs.

To alleviate the bias in fault diagnosis caused by imbalanced training data, researchers have conducted extensive research at both the data level and model level. At the data level, data generation methods are mainly used to balance the number of samples from different classes participating in iterative training. Reference [18] used the synthetic minority oversampling technique to relieve the data imbalance problem. Reference [19] introduced a weakly supervised learning framework for oversampling to generate minority samples. Reference [20] adopted a GAN to generate a small number of samples to mitigate the influence of imbalanced data on WT fault diagnosis. References [21] and [22] further developed different GAN-based models, attempting to improve the accuracy of WT fault diagnosis with insufficient fault samples by enhancing the similarity between generated data and real data. Experimental results have demonstrated the ef-

fectiveness of these models in mitigating classification bias caused by imbalanced data through data generation. However, models trained on limited real samples for data generation are susceptible to overfitting, which often leads to low diversity in the generated data and an inadequate representation of the underlying true distribution. Consequently, when such data are combined with real data for fault diagnosis, the reliability of the diagnostic results may be compromised.

Research at the model level focuses on adjusting the loss function to improve the model learning ability for imbalanced data. Reference [23] proposed a CNN-based WT fault diagnosis method, which combined cross-entropy (CE) loss and Dice loss to increase the attention to minority samples during model training. Reference [24] introduced a multi-learner neural network equipped with a label-distribution-aware margin (LDAM) loss to improve diagnostic accuracy under imbalanced data. Reference [25] developed an adaptive weighted loss function to address data imbalance in icing detection of WT blade, combined with a spatiotemporal attention model (SAM) for improved real-world performance. Reference [26] employed a spatiotemporal multi-scale neural network (STMNN) and replaced the standard CE loss with focal loss to mitigate the dominance of majority samples. Reference [27] added a focal loss term to the cost function of reinforcement learning to reduce the contribution of easily classified samples to the total loss, and experimental results proved the effectiveness of this improvement. The essence of loss function adjustment lies in reinforcing attention to minority samples by tuning the weights during the model training process, aiming to mitigate the bias toward certain categories. Compared with data generation methods, the overfitting risk associated with introducing additional generated samples is mitigated. As the training data remain unchanged, minimal computational costs are incurred and faster training is enabled. Inspired by these studies, this paper proposes a spatiotemporal feature extraction framework for WT fault diagnosis, especially under imbalanced data.

To comprehensively explore the hidden information in SCADA data, we develop an integrated multivariate time series (IMTS) based on sliding window sampling. Unlike variable-wise processing that ignores spatial correlations [8] and graph-based methods requiring predefined adjacency matrices [28], IMTS explicitly preserves the physical spatial topology of sensor networks through its matrix structure, enabling effective identification of subsystem-level fault characteristics by deep learning models. Its dynamic sliding window strategy maintains complete fault evolution continuity while diversifying training samples through contextual variations. Notably, the dual-label IMTS synergistically combines global operation status labels with local fault type labels, simultaneously exploiting spatiotemporal feature fusion and mitigating class imbalance, thereby providing structured input for subsequent feature extraction. In this paper, the majority class presents the normal operation states of WT. The minority class presents faulty operation states of WT. A hierarchical model named multi-scale convolutional neural network (MCNN)-LSTM is proposed for multiple classes of WT fault

diagnosis based on IMTS. Compared with single-scale CNN-LSTM model limited to fixed receptive fields [15], the proposed MCNN-LSTM model uses an MCNN to extract the spatial patterns from IMTSs at multiple scales, generating prior probability estimates of overall faulty operation states of WT. Concurrently, a stacked LSTM network is used to learn the fusion information from individual sample within the IMTS for precise state diagnose of the target sample. Then, the prior probability estimates are used to correct the diagnostic results to enhance the reliability. Whereas weighted cross-entropy (WCE) loss and focal loss rely on static weighting schemes, we develop an improved weighted cross-entropy (IWCE) loss function that incorporates kernel density estimation (KDE)-based sample probabilities as regularization terms to dynamically adjust decision boundaries, thereby enhancing model performance under imbalanced data.

The rest of this paper is organized as follows. Section II provides SCADA data description and WT fault diagnosis flow. Section III introduces the proposed MCNN-LSTM model. Experimental verifications are presented in Section IV. Finally, Section V draws the conclusions.

II. SCADA DATA DESCRIPTION AND WT FAULT DIAGNOSIS FLOW

A. SCADA Data Description

The SCADA system for a WT employs sensors distributed throughout the unit to collect hundreds of operation parameters. Supplementary Material A Fig. SA1 shows the common sensor installation locations in WTs. Supplementary Material A Fig. SA2 presents the exported condition monitoring data, operation logs, and fault logs from the SCADA system. Collectively, these three categories provide the research foundation for anomaly detection and fault diagnosis in WTs.

Not all variables in SCADA data are relevant to fault diagnosis. Redundant variables hinder feature extraction and increase computational costs, making the selection of discriminative features essential. Component failures in WTs are typically preceded by gradual degradation, during which short-term temporal dependencies such as evolving trends and transient patterns in the state data provide critical indicators of incipient anomalies. Moreover, strong coupling exists between internal components. A fault in one component often induces coordinated changes across multiple monitored variables. For example, a gearbox bearing fault may simultaneously affect oil temperature, bearing temperatures, rotational speed, and current. Extracting these cross-variable spatial correlations can substantially improve diagnostic accuracy. Therefore, jointly leveraging temporal dependencies within variables and spatial correlations among variables in SCADA data is a key strategy for enhancing the performance of anomaly detection and fault diagnosis in WTs.

B. WT Fault Diagnosis Flow

The WT fault diagnosis flow under imbalanced data is shown in Supplementary Material A Fig. SA3, which consists of four key stages. First, multi-source SCADA data undergo feature selection and preprocessing to construct an

IMTS sample set for model training, along with fault-specific sample set for distribution estimation. Each IMTS organizes continuous multi-variable monitoring readings into a two-dimensional (2D) matrix format, preserving both temporal continuity and inter-variable dependencies. Each timestamp within an IMTS is assigned a local label reflecting its operation state, while the entire sequence carries a global label indicating whether any fault occurs during the monitoring period. Subsequently, an MCNN-LSTM model is constructed for multi-class fault diagnosis. Within the model, MCNN and stacked LSTM network interact by dynamically aligning spatial and temporal representations, enabling deep fusion of complementary diagnostic information, where spatial patterns inform temporal analysis, and temporal patterns refine spatial feature interpretation. Next, KDE-based sample probabilities are employed to approximate the conditional probability density functions (PDFs) of individual samples under given operation conditions, revealing the underlying fault probability distribution. Finally, the KDE-based sample probabilities are incorporated as external evidence into the WCE loss to form the IWCE loss. This formulation enhances fault diagnosis performance when dealing with imbalanced data.

III. PROPOSED MCNN-LSTM MODEL

A. Data Preprocessing

Data preprocessing plays a critical role in enhancing model performance, mitigating overfitting risks, and reducing computational costs. The data preprocessing workflow for the spatiotemporal feature extraction framework is illustrated in Supplementary Material A Fig. SA4. Details of data preprocess and feature selection are shown in Supplementary Material B.

B. IMTS Sample Construction

Assume that m variables are determined for model training after data preprocessing and feature selection. Then, the running logs provided by the SCADA system are combined to assign labels to each piece of monitoring data using the timestamp as the index. These labels are called individual labels. Within these logs, different faults are uniquely encoded as integers c ($c = 1, 2, \dots, K$), while the data without recorded faults in the logs are considered normal and encoded as 0. The raw samples obtained through continuous sampling from the SCADA system are referred to as individual samples to distinguish them from IMTS samples. A sliding window with a fixed length L is used to perform sliding sampling on the labeled individual samples at a fixed rate Δv , thereby generating a series of IMTS samples. In addition, a logical “OR” operation is performed on the individual labels within each IMTS sample to determine whether it contains a fault. If a fault is contained, the IMTS is marked as 1 and regarded as an anomaly sample; otherwise, it is marked as 0 and regarded as a normal sample. These labels assigned to IMTS samples are referred to as global labels. The process of IMTS sample construction is shown in Fig. 1, where X_i is an $L \times m$ matrix composed of L consecutive individual samples; Y_i records the state of the L individual samples within

X_i ; f_i is the global label of the IMTS sample, indicating whether the IMTS contains the fault ($f_i=1$) or not ($f_i=0$); and V_n ($n=1,2,\dots,d$) is the n -dimensional monitoring variable.

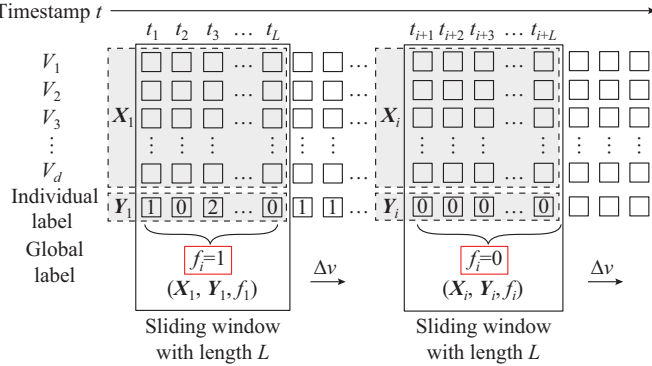


Fig. 1. Process of IMTS sample construction.

The supervised IMTS dataset D for WT fault diagnosis is defined as:

$$D = \{(X_i, Y_i, f_i)\}_{i=1}^N \quad (1a)$$

$$X_i = \begin{bmatrix} x_{i,1}^1 & x_{i,1}^2 & \dots & x_{i,1}^m \\ x_{i,2}^1 & x_{i,2}^2 & \dots & x_{i,2}^m \\ \vdots & \vdots & \ddots & \vdots \\ x_{i,l}^1 & x_{i,l}^2 & \dots & x_{i,l}^m \\ \vdots & \vdots & \ddots & \vdots \\ x_{i,L}^1 & x_{i,L}^2 & \dots & x_{i,L}^m \end{bmatrix} \in \mathbb{R}^{L \times m} \quad (1b)$$

$$Y_i = \begin{bmatrix} y_{i,1} \\ y_{i,2} \\ \vdots \\ y_{i,l} \\ \vdots \\ y_{i,L} \end{bmatrix} \quad (1c)$$

$$f_i = \begin{cases} 0 & \sum_{t=1}^L y_{i,t} = 0 \\ 1 & \sum_{t=1}^L y_{i,t} \geq 1 \end{cases} \quad (1d)$$

where N is the number of IMTSs; $y_{i,t}$ is the local fault label at time t in the i^{th} IMTS; and $x_{i,l}^m$ is the m^{th} variable of the l^{th} ($1 \leq l \leq L$) individual sample in the i^{th} IMTS.

The IMTS provides three key advantages. First, its matrix-based structure explicitly preserves the physical spatial topology of the sensor network. This structural inductive bias enables the proposed MCNN-LSTM model to learn physically meaningful spatiotemporal correlations, which are fundamental for accurate subsystem-level fault characterization. Second, the dynamic sliding window not only maintains complete state transition information before and after fault occurrences, but also significantly expands the effective training dataset by combining identical fault samples with varying contextual information, thereby enhancing the capability of the proposed MCNN-LSTM model to generalize across different fault scenarios. Third, the dual-label IMTS is designed

to fully exploit spatiotemporal characteristics embedded in SCADA data, combining X_i to train the MCNN and stacked LSTM network, respectively. Specifically, the global label guides the proposed MCNN-LSTM model to learn system-level spatiotemporal correlations, while the local label focuses on extracting detailed fault related features. This dual-supervision strategy not only promotes more comprehensive utilization of spatiotemporal information but also alleviates the impact of sample imbalance at the data level.

C. Architecture of Proposed MCNN-LSTM Model

1) Spatial Feature Learning Module of MCNN

CNN is a type of neural network structure designed to simulate the visual process of mammals. Its local perception and weight-sharing mechanisms make it effective in feature extraction from 2D data. Therefore, it is widely used in object detection and computer vision. However, various faults exhibit different manifestations, and different faults may need to be characterized at different scales [29]. Traditional CNN uses fixed-size filters for feature extraction, which may affect the discriminative representation of different faults because it cannot fully exploit the hidden information in the data. Based on [30] and [31], an MCNN is proposed to learn distinctive features from IMTS.

MCNN introduces z filters with different sizes to perform independent feature extraction on IMTS in different branches, allowing the MCNN to naturally learn multi-scale features from coarse to fine. Each branch of the MCNN contains s convolution units, including multiple operations such as convolution, batch normalization, activation, and average pooling, which can be expressed as:

$$X_{i,j}^a = \theta(\text{AveP}(\text{BN}(\text{Conv}(X_i^a, k_j)))) \quad j=1,2,\dots,z, a=1,2,\dots,s \quad (2)$$

where $\text{Conv}(\cdot)$ represents the convolution operation; $\text{BN}(\cdot)$ represents the batch normalization operation; $\text{AveP}(\cdot)$ represents the average pooling operation; $\theta(\cdot)$ represents the ReLU activation function; X_i^a is the 2D input of the a^{th} feature extraction unit; $X_{i,j}^a$ is the output processed by the a^{th} convolution unit in the j^{th} branch; and $k_j = 2(j-1)+1$, which is the size of the 2D convolution kernel in the j^{th} branch. The 2D features obtained through s convolution units can be converted into 1D vectors through the flattening operation. Subsequently, a more informative and effective spatial feature representation can be obtained through multi-scale spatial feature concatenation and fully-connected layer processing.

$$S_i = \text{Fc}(X_{i,1}^s \oplus X_{i,2}^s \oplus \dots \oplus X_{i,z}^s) \quad (3)$$

where $\text{Fc}(\cdot)$ represents a fully-connected operation; $\oplus$ represents the concatenation operation of vectors; $X_{i,j}^s$ ($j=1,2,\dots,z$) represents the features extracted by the j^{th} branch through the s^{th} convolution unit; and S_i is the spatial feature obtained through the MCNN, which is a $1 \times m$ vector.

As can be observed from the illustration of multi-scale spatial feature extraction based on the MCNN, the interaction between the MCNN and stacked LSTM network is reflected in two aspects. On the one hand, the multi-scale S_i extracted by the MCNN will be fused with individual samples in the IMTS by a fully-connected layer and then fed to the stacked LSTM network for further learning of spatiotem-

poral features. On the other hand, the binary logit output $[z_0, z_1]$ can be obtained by performing a fully-connected layer transformation on $\mathbf{S}_p$, and the failure probability of IMTS P_i can be obtained through the softmax function as:

$$P_i(\hat{f}_i=1) = \frac{\exp(z_1)}{\exp(z_1) + \exp(z_0)} \quad (4)$$

where $\exp(\cdot)$ represents the exponential function; $\hat{f}_i$ is the predicted external label value of the i^{th} IMTS; and z_0 and z_1 are the logit scores for normal and abnormal IMTS, respectively. The diagnosis of IMTS by the MCNN is a binary classification problem. Therefore, binary cross entropy is used as the loss function to ensure diagnostic reliability.

$$L(f, \hat{f}) = -\frac{1}{N} \sum_{i=1}^N [f_i \lg(\hat{f}_i) + (1-f_i) \lg(1-\hat{f}_i)] \quad (5)$$

where $L(f, \hat{f})$ is the loss function of the MCNN.

2) Spatiotemporal Feature Learning Based on Stacked LSTM Network

The stacked LSTM network is chosen for its advantage in spatiotemporal feature learning, owing to its ability to capture long-term dependencies and overcome the vanishing gradient problem via gate control. The detailed structure of an LSTM unit and the functions of its internal gates can be found in Supplementary Material C. To further enhance feature extraction for complex spatiotemporal patterns, a stacked LSTM network is adopted, which enables the learning of hierarchical representations, where lower layers capture short-term fluctuations and higher layers integrate these into long-term trends. The inputs to the first LSTM layer are derived through an information fusion process described as:

$$\mathbf{x}_{i,t}^* = Fc(\mathbf{S}_i \oplus \mathbf{x}_{i,t}) \quad (6)$$

where $\mathbf{x}_{i,t}^*$ is the input of the LSTM, which is obtained by fusing $\mathbf{S}_i$ and the individual sample at time t in the i^{th} IMTS by the fully-connected layer; and $\mathbf{x}_{i,t}$ is the raw monitoring sample at time t in the i^{th} IMTS. The spatiotemporal features extracted through multiple LSTM hidden layers are transformed by the fully-connected layer. Then, the probability values of the last individual sample $\mathbf{x}_{i,l}$ in the i^{th} IMTS belonging to different categories can be derived as:

$$p_{i,l}^c = \frac{e^{z_c}}{\sum_{j=0}^K e^{z_j}} \quad c = 1, 2, \dots, K \quad (7)$$

where $p_{i,l}^c$ is the probability that the l^{th} individual sample in the i^{th} IMTS belongs to class c ; z_c is the logit value output of class c by the fully-connected layer; and K is the total number of categories for class c . In the proposed MCNN-LSTM model, the probability output by the stacked LSTM network is not directly used as the final basis for fault diagnosis. Instead, the empirical weighting method is used to correct the fault probability predicted by the stacked LSTM network. The overall fault probability P_i predicted by the MCNN is used as the correction reference.

$$\hat{p}_{i,l}^c = (1-\omega)p_{i,l}^c + \omega p_{i,l}^{c,tar} \quad c = 0, 1, \dots, K \quad (8a)$$

$$p_{i,l}^{c,tar} = \begin{cases} P_i & c = \arg \max c(p_{i,l}^c) \\ \frac{1-P_i}{K-1} & \text{else} \end{cases} \quad (8b)$$

where $\hat{p}_{i,l}^c$ is the corrected probability that the individual sample in the i^{th} IMTS is diagnosed as fault class c ; $p_{i,l}^{c,tar}$ is the expected probability that the individual sample is diagnosed as different categories in the i^{th} IMTS; ω is the probability fusion weight, whose value reflects the impact of the overall fault probability of IMTS predicted by the MCNN on the diagnostic results; and $\arg \max c(p_{i,l}^c)$ denotes the selection of c that maximizes the value of $p_{i,l}^c$. In the probability correction method, the fault diagnosis result of an individual sample is constrained by the fault diagnosis result of the IMTS, which reduces the influence of extreme individual cases on the diagnostic results and helps improve the accuracy of fault diagnosis.

3) Construction of IWCE Loss

WCE loss is designed to handle imbalanced data in multi-class classification problems. WCE loss assigns class weights inversely proportional to their frequency, enabling the proposed MCNN-LSTM model to focus on the minority class and improve the performance under imbalanced data. The minority class is not necessarily hard, and the majority class may contain hard-to-classify samples. However, the WCE loss does not distinguish between easy and hard-to-classify samples, and only considers the frequency of each category, which may still lead to insufficient classification accuracy in complex situations.

To overcome the limitations of WCE loss and ensure the accuracy of fault diagnosis under imbalanced data, IWCE loss is proposed, which enhances the attention of the proposed MCNN-LSTM model to a minority class and hard-to-classify samples by adding an adjustment term based on external observations to WCE loss. IWCE loss is defined as:

$$\begin{cases} L_{IWCE} = -\sum_{i=1}^{N_{all}} \sum_{c=0}^{K-1} (\alpha_c y_{i,c} \lg(\hat{p}_{i,l}^c) + \lambda q_{i,c}) \\ \alpha_c = \frac{N_{all}}{N_c} \end{cases} \quad (9)$$

where α_c is the weighting factor for class c ; N_{all} is the total number of individual samples; N_c is the number of individual samples of class c ; $q_{i,c}$ is the external reference probability for the i^{th} individual sample belonging to class c ; and λ is the adjustment coefficient used to control the influence of $q_{i,c}$ on the IWCE loss. The main reason why some samples are difficult to classify is that the probabilities predicted by the proposed MCNN-LSTM model for different classes are very small or follow a uniform distribution, which results in low confidence in classifying these samples. To address this issue, the IWCE loss introduces an adjustment term derived from external observations. This design provides the proposed MCNN-LSTM model with valuable reference information, enhancing its ability to assess the true class likelihood of each sample. While retaining the fundamental structure of WCE loss, which inherently emphasizes the minority class, IWCE loss further incorporates sample-level distributional awareness. By integrating KDE-based sample probabilities as a form of contextual knowledge, the proposed MCNN-LSTM model becomes better equipped to handle ambiguous and hard-to-classify samples. As a result, IWCE loss contributes to improved diagnostic performance under data imbalance.

In this paper, the additional external reference is the conditional probability that the predicted sample belongs to different classes obtained from external observations. The KDE-based sample probabilities are used to estimate the value of the adjustment factor, which can also be used to estimate the PDF without prior knowledge or distribution assumptions. The PDFs of various classes based on KDE-based sample probabilities can be expressed as (10), then $q_{i,c} = f_c(\mathbf{x}_{i,l})$, $c = 0, 1, \dots, K$.

$$f_c(\mathbf{x}) = \frac{1}{N_c h^d} \sum_{\mathbf{x}_j^c \in D_c, j=1}^{N_c} \text{Kel}\left(\frac{\mathbf{x} - \mathbf{x}_j^c}{h}\right) \quad (10)$$

where $f_c(\mathbf{x})$ is the PDF of class c ; $\mathbf{x}_j^c$ is the j^{th} sample in the sample set D_c of class c ; $\text{Kel}(\cdot)$ is the Gaussian kernel function; h is the bandwidth; and $\mathbf{x}$ is the given conditional sample.

IV. EXPERIMENTAL VERIFICATION

A. Metric

To evaluate the fault diagnosis performance of the proposed MCNN-LSTM model under imbalanced data, several commonly used metrics are employed, and their detailed illustrations are presented in Supplementary Material A Table SAI. Class-specific precision and recall are computed to assess discriminative capability for each class. The F1-score, which balances precision and recall, serves as a performance indicator for individual classes. To evaluate the overall performance across all K categories, the macro F1-score (Macro-F1) and macro recall (Macro-recall) are calculated by averaging class-wise F1-scores and recall values, respectively, thereby mitigating bias toward the majority class. The overall diagnostic accuracy measures the proportion of correctly classified samples across all categories. The mean of class-wise recall (G-mean) further accounts for inter-class performance imbalance, thereby enhancing the robustness in imbalanced settings. Finally, the Kappa coefficient (denoted as Kappa) quantifies classification agreement beyond random chance. Kappa isolates statistically meaningful performance from coincidental correctness. Together, these metrics address both per-class granularity and overall generalization, which are critical for validating fault diagnosis performance of the proposed MCNN-LSTM model under imbalanced data.

B. Data Description and Experimental Setup

The effectiveness of the proposed MCNN-LSTM model is verified using data collected from a 1.5 MW doubly-fed induction generator (DFIG) from April 1, 2018 to January 31, 2019. The SCADA system of the wind farm records condition monitoring variables at 5-min intervals, along with operation and fault logs. After data preprocessing and labeling, 61247 valid individual samples are obtained. According to the fault handling manual provided by the manufacturer, faults are grouped into 5 major categories and 29 subcategories. Only major categories are considered in this paper. Therefore, together with normal operation data, these data are divided into six categories in this paper. Category descriptions, labels, and counts are provided in Supplementary

Material A Table SAII.

An empirical threshold of 0.5 is applied for feature selection, retaining 36 out of the original 103 variables, as detailed in Supplementary Material A Table SAIII. A sliding window with a length of 288 (corresponding to 24 hours) under $\Delta v = 1$ is applied, yielding 52485 normal and 8745 abnormal IMTS samples. To mitigate class imbalance, a balanced dataset is formed by randomly selecting 8475 normal IMTS samples and combining them with all 8475 abnormal IMTS samples. The balanced dataset undergoes stratified five-fold cross-validation with dual preservation criteria: global label balance, which ensures equal normal-anomaly ratios across folds, while fault type consistency maintains the original imbalanced distribution of five diagnostic categories. The fault label for each abnormal IMTS sample is assigned based on the timestamp of its last constituent sample, ensuring that rare faults are adequately represented and preventing validation bias. The design of unit-step sliding window preserves temporal continuity between adjacent IMTS samples, enhancing model capacity to learn progressive failure patterns through spatiotemporal coherence.

The MCNN employs five parallel branches for multi-scale spatial feature extraction, each consisting of two convolutional units with kernel sizes of 1, 3, 5, 7, and 9, respectively. The stacked LSTM network comprises three hidden layers with 288, 100, and 50 neurons, respectively. The overall architecture is summarized in Supplementary Material A Table SAIV, with hyperparameters selected via empirical testing and grid search. During training, the Adam optimizer is used with a learning rate of 0.0001, a weight decay of 0.0002, and a batch size of 32. The training runs for up to 100 epochs with early stopping. $\omega = 0.4$ and $\lambda = 0.3$ are selected via grid search. All experiments are conducted on an Intel dual-core i7-11700 CPU (2.50 GHz) using Python 3.9.7 and TensorFlow 2.10.0 on a 64-bit Windows system.

To systematically validate the effectiveness of the proposed MCNN-LSTM model, four experimental schemes are designed and evaluated:

- 1) Scheme 1: probability correction is not adopted, and the WCE loss is adopted.
- 2) Scheme 2: probability correction and the WCE loss are adopted.
- 3) Scheme 3: probability correction is not adopted and the IWCE loss is adopted.
- 4) Scheme 4: probability correction and the IWCE loss are adopted.

Table I summarizes the precision and recall values for each class in four schemes. The majority class refers to Label 0, which represents the normal operating state of WT. The minority class refers to Labels 1-5, which represents five major fault states of WT. Label 0 achieves consistently high performance in all schemes. However, Label 4 performs poorly in Scheme 1. For example, the precision and recall for Label 4 are only 0.291 and 0.642, respectively. Scheme 2 incorporates probability calibration alongside WCE loss, yielding modest performance gains: the precision for Label 4 increases to 0.339, and the recall rises to 0.663. In Scheme 3, the WCE loss is replaced with the IWCE loss,

leading to a substantial improvement in the minority class. The precision for Label 4 increases to 0.662 and the recall for Label 4 improves to 0.881. When probability calibration and IWCE loss are jointly applied in Scheme 4, the proposed MCNN-LSTM model achieves the best performance. The precision for Label 4 further increases to 0.857, and the

recall for Label 4 improves to 0.884. These results demonstrate that combining calibrated confidence estimation with an imbalance-aware loss significantly enhances the discriminative capability for rare fault categories while preserving the spatiotemporal capacity of the proposed MCNN-LSTM model.

TABLE I
PRECISION AND RECALL VALUES FOR EACH CLASS IN FOUR SCHEMES

Class	Scheme 1		Scheme 2		Scheme 3		Scheme 4	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Label 0	0.992 ± 0.001	0.911 ± 0.004	0.994 ± 0.001	0.929 ± 0.004	0.997 ± 0.001	0.973 ± 0.004	0.997 ± 0.001	0.987 ± 0.002
Label 1	0.261 ± 0.021	0.668 ± 0.043	0.303 ± 0.009	0.689 ± 0.032	0.585 ± 0.041	0.898 ± 0.024	0.707 ± 0.044	0.915 ± 0.030
Label 2	0.165 ± 0.009	0.632 ± 0.030	0.218 ± 0.024	0.664 ± 0.032	0.537 ± 0.052	0.911 ± 0.030	0.702 ± 0.028	0.920 ± 0.044
Label 3	0.194 ± 0.028	0.635 ± 0.044	0.257 ± 0.019	0.686 ± 0.032	0.567 ± 0.055	0.887 ± 0.034	0.752 ± 0.076	0.875 ± 0.047
Label 4	0.291 ± 0.057	0.642 ± 0.091	0.339 ± 0.038	0.663 ± 0.054	0.662 ± 0.017	0.881 ± 0.026	0.857 ± 0.020	0.884 ± 0.051
Label 5	0.201 ± 0.023	0.484 ± 0.020	0.279 ± 0.027	0.671 ± 0.038	0.515 ± 0.013	0.858 ± 0.015	0.758 ± 0.058	0.864 ± 0.024

Figure 2 presents comprehensive metrics for the four schemes. From Scheme 1 to Scheme 4, systematic improvements are observed. Scheme 1, serving as the baseline, yields suboptimal performance, with Macro-F1 score, G-mean score, and Kappa score of 0.427, 0.649, and 0.354, respectively, reflecting the detrimental impact of class imbalance. Scheme 2 introduces probability calibration, elevating these metrics to 0.487, 0.711, and 0.428, confirming that confidence recalibration mitigates misclassification bias. In Scheme 3, Macro-F1 score surges to 0.765, G-mean score increases to 0.863, and Kappa score surges to 0.715. This verifies the effectiveness of external observation-based tuning in enhancing attention to the minority class. In Scheme 4, Macro-F1 score reaches 0.845, G-mean score stabilizes at 0.906, and Kappa score reaches 0.825. These results confirm that probability calibration refines decision boundaries, while the IWCE loss enhances feature learning, providing a robust solution for diagnosing imbalanced multivariate time-series data in WT fault diagnosis.

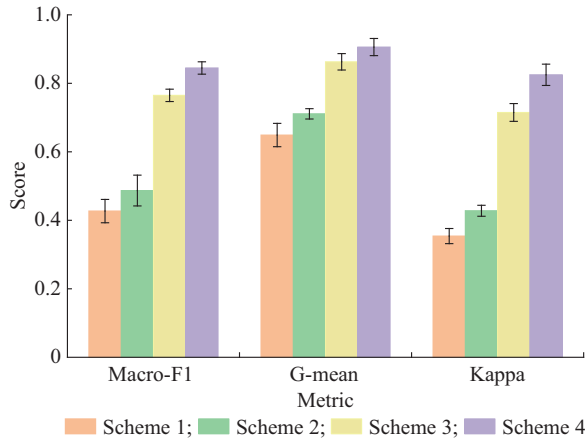


Fig. 2. Comprehensive metrics for four schemes.

C. Parameter Settings

1) Weight Setting

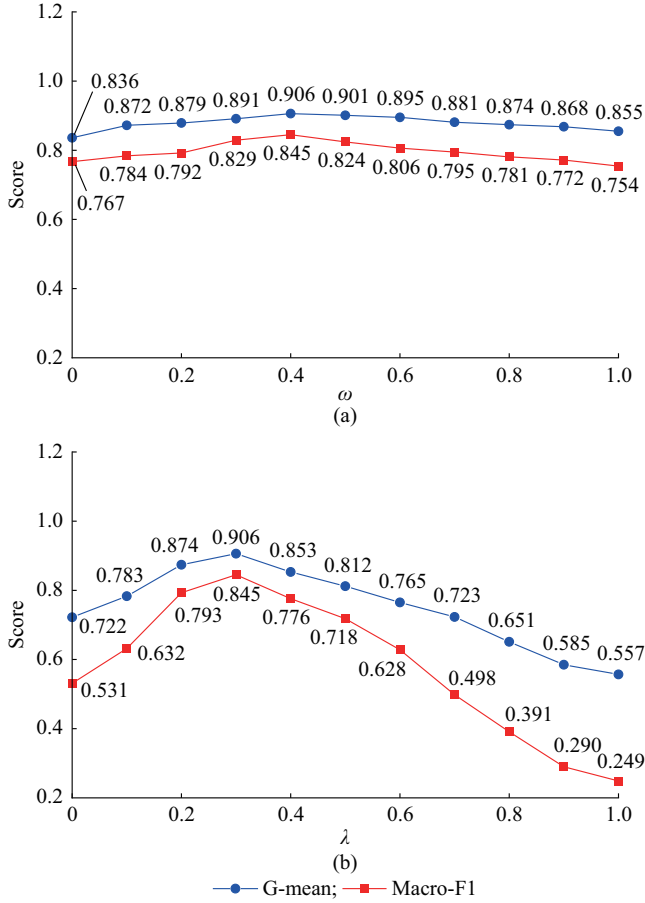
To assess the impact of ω and λ , a sensitivity analysis is

conducted based on Scheme 4. As shown in Fig. 3, both Macro-F1 and G-mean scores exhibit non-monotonic dependence on these hyperparameters, with peak performance achieved when $\omega=0.4$ and $\lambda=0.3$. Notably, λ exerts a more pronounced effect on model performance than ω , underscoring the key role of loss function design for class imbalance. The results of the parameter sensitivity analysis, as shown in Table II, confirm that this configuration not only yields optimal accuracy but also demonstrates the robustness to minor parameter perturbations, as evidenced by stable G-mean score and low standard deviations. These results justify the selected parameter values and underscore the necessity of jointly tuning the confidence calibration and loss weighting in fault diagnosis. Detailed experimental procedures and extended results are provided in Supplementary Material D.

2) Performance Evaluation of IWCE Loss

To evaluate the effectiveness of the IWCE loss, five loss functions, including CE loss, WCE loss, IWCE loss, focal loss, and Dice loss [32], are used to train the proposed MCNN-LSTM model without probability correction. λ of IWCE loss is set to be 0.3, the smoothing coefficient of Dice loss is set to be 1, and the focus factor γ and balance factor α of the focal loss are set to be 2 and 0.25, respectively [26].

Figure 4 shows the model performance evaluation based on different loss functions across multiple metrics. In Fig. 4(a), CE loss performs poorly in imbalanced settings, yielding low Macro-F1 score of 0.394, Macro-recall score of 0.582, and G-mean score of 0.565, despite a high overall accuracy of 0.896, indicating majority class bias. In comparison, the IWCE loss achieves the best performance, with Macro-recall score of 0.908, Macro-F1 score of 0.845, and G-mean score of 0.906, outperforming the second-best focal loss by 1.2%, 0.36%, and 1.2%, respectively. Although focal loss slightly outperforms IWCE loss in Kappa (0.831 versus 0.825) and accuracy (0.984 versus 0.983), IWCE loss delivers more consistent improvements across the metrics.

Fig. 3. Model performance based on different hyperparameters. (a) w . (b) λ .TABLE II
RESULTS OF PARAMETER SENSITIVITY ANALYSIS

Parameter	Macro-F1 score	G-mean score
$w=0.3$	0.829 ± 0.04	0.902 ± 0.03
$w=0.4$	0.845 ± 0.03	0.906 ± 0.02
$w=0.5$	0.824 ± 0.03	0.901 ± 0.02
$\lambda=0.2$	0.794 ± 0.01	0.895 ± 0.01
$\lambda=0.3$	0.845 ± 0.03	0.906 ± 0.02
$\lambda=0.4$	0.756 ± 0.03	0.868 ± 0.02

The radar chart in Fig. 4(b) further illustrates that IWCE loss and focal loss outperform other loss functions, with IWCE loss achieving a marginally larger normalized area, reflecting more uniform improvement across all evaluation dimensions. This balanced enhancement arises from the integration of adaptive category weighting and external observation mechanisms, prioritizing a few samples and optimizing feature learning. These characteristics enable the IWCE loss to achieve excellent diagnostic performance in imbalanced multivariate time-series classification tasks.

3) Sliding Window and Model Architecture

The size of the sliding window affects both the number of generated multivariate time-series samples and the capability of the proposed MCNN-LSTM model to capture spatiotemporal features, while the model architecture is empirically optimized through extensive experiments.

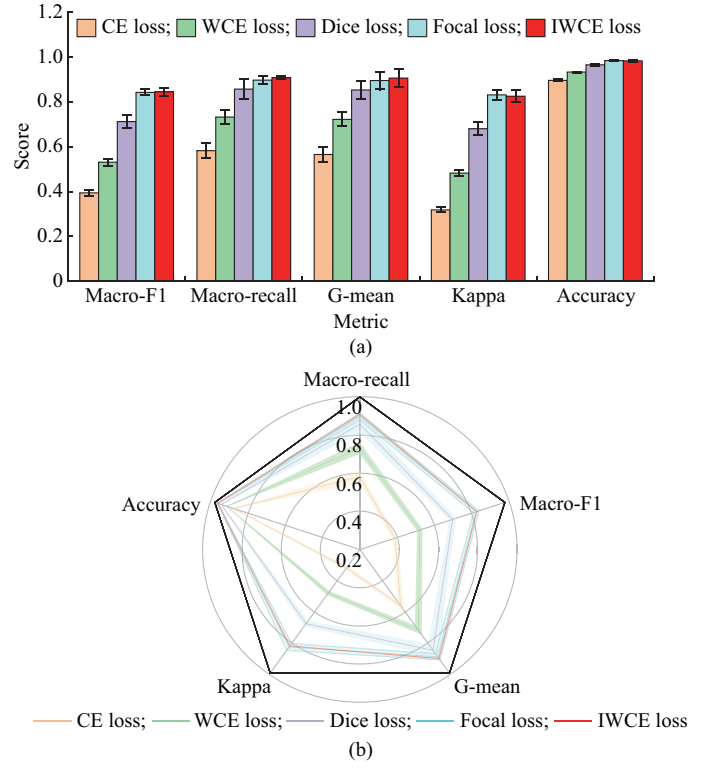


Fig. 4. Model performance evaluation based on different loss functions across multiple metrics. (a) Bar chart of metrics. (b) Radar chart of metrics.

As these aspects represent implementation-level design choices rather than the core contributions of this paper, and detailed comparative and ablation studies are provided in Supplementary Material E.

D. Verification by Comparative Experiment

To validate the diagnostic advantages of the proposed MCNN-LSTM model under imbalanced data, comparative experiments are conducted at both data level and model level.

1) Data-level Comparison

Four data augmentation models are employed to transform the original imbalanced data into balanced data.

1) Synthetic minority over-sampling technique (SMOTE) is a classic linear interpolation-based oversampling model, serving as a baseline for comparison. The k -neighbors is set to be 5.

2) Borderline-SMOTE is an advanced oversampling model derived from SMOTE, which generates synthetic samples near the decision boundary to enhance detection of incipient faults [33]. The k -neighbors is set to be 5.

3) SMOTE + ENN is a hybrid resampling model combining SMOTE and edited nearest neighbour (ENN) for noise filtering [34]. The k -neighbors for SMOTE is set to be 5, and the n -neighbors for ENN is set to be 3.

4) Conditional generative adversarial network (CGAN) model can learn to generate realistic samples based on category labels [19]. In the comparative experiment, the generator adopts a three-layer perceptron structure with a learning rate of 0.001.

All comparative experiments adopt an identical evaluation

framework, with the model architecture, IWCE loss, and IMTS sample format kept consistent across all tests. All models are tested on the same raw imbalanced testing data. The only difference lies in the training data composition: the proposed MCNN-LSTM model uses the raw imbalanced training data, whereas the other models use balanced training data reconstructed by the above models. To accommodate the dual-label IMTS, a category-oriented stratified oversampling strategy is adopted. The minority classes under per-timestamp individual labels are augmented to 500 samples each. Each synthetic fault sample is injected by replacing the final time point of a randomly selected normal IMTS sample. To ensure temporal consistency, the candidate carrier IMTS is screened by verifying that the absolute deviation of key operation characteristics (e.g., wind speed, shaft speed, and gearbox oil temperature) is below 10%. Finally, the generated data are mixed with the original training data to form a new training set.

Table III reports the metrics and computational cost for all models. The results show that the proposed MCNN-LSTM model achieves competitive performance without any data augmentation. Although the overall scores are slightly lower than those achieved by Borderline-SMOTE, SMOTE + ENN, and CGAN, the gap remains small, which is less than 2.5% compared with the best-performing CGAN in key metrics such as Macro-F1 and G-mean. Moreover, it clearly outperforms SMOTE, which lacks both noise filtering and boundary-aware sample generation. Figure 5 presents the class-wise recall results for each class. All models achieve similar performance on the majority class, while significant differences are observed in the minority class. The proposed MCNN-LSTM model maintains good performance across Labels 1-5. For example, it matches CGAN on Label 1 with recall of 0.957 and achieves the same recall as SMOTE on Label 4.

TABLE III
METRICS AND COMPUTATIONAL COST FOR ALL MODELS

Model	Score				Accuracy	Computational cost (s)
	Macro-F1	Macro-rec	G-mean	Kappa		
SMOTE	0.871	0.933	0.933	0.864	0.987	2204.3
Borderline-SMOTE	0.883	0.954	0.953	0.871	0.988	2265.2
SMOTE + ENN	0.892	0.954	0.954	0.872	0.988	2273.0
CGAN	0.907	0.966	0.966	0.889	0.990	2591.6
MCNN-LSTM	0.884	0.943	0.942	0.867	0.988	1632.7

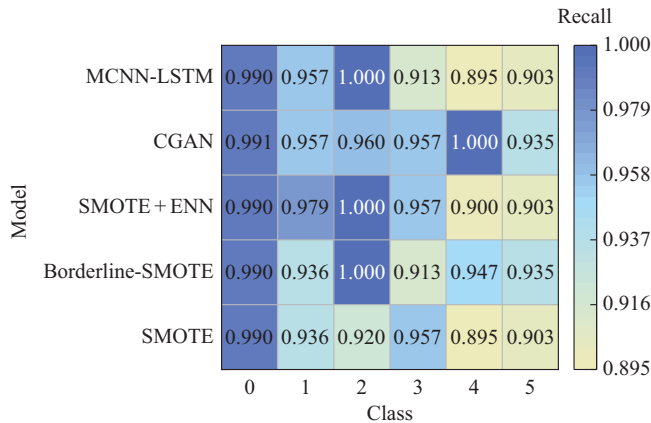


Fig. 5. Class-wise recall results for each class.

Comprehensive comparisons demonstrate that, while data augmentation models can enhance overall model performance under imbalanced data, their effectiveness is inherently constrained. SMOTE variants rely on interpolation accuracy for moderately sized classes, while CGAN depends on pattern generalization for extremely rare classes, neither achieving universal optimization across all classes. In contrast, the proposed MCNN-LSTM model handles data imbalance implicitly through the optimization of training sample, model architecture, and loss function design. It attains 97.5% of the key diagnostic performance achieved by the CGAN, yet incurs only 63% of the corresponding computa-

tional cost, thereby underscoring its practical advantage.

2) Model-level Comparison

Seven representative deep learning-based models for WT fault diagnosis are selected for model-level comparison, all of which extract spatiotemporal features from SCADA data.

1) MKFCNN is a multi-kernel fusion CNN-based model that uses SCADA data for WT fault diagnosis [35].

2) CNN-GRU is a WT fault diagnosis model based on spatiotemporal feature fusion of SCADA data using CNN and GRU [36].

3) SCNN-LSTM is a series network-based model for WT fault diagnosis constructed by using CNN and LSTM [37].

4) PCNN-LSTM is a parallel-network-based model using CNN and LSTM, which determines WT faults by parallel diagnosis from both branches [15].

5) STMNN is a STMNN model for WT fault diagnosis under imbalanced data [26].

6) EMCAN is a fault diagnosis model based on the combination of an ensemble multi-scale convolutional attention network and a weighted voting strategy [38].

7) STA-SAW is a WT fault diagnosis model that combines a spatiotemporal attention network with an adaptive weighted loss function [25].

All models are trained and tested on the same datasets used in the prior experiments. Hyperparameters are tuned via grid search within the original ranges in literature. Since the first four models of the above seven models are not specifically designed for imbalanced data, their loss functions are

replaced with $\alpha=0.25$ and $\gamma=2.0$ to ensure a fair comparison. The remaining three models retain their published imbalance-handling strategies. Core training settings, including learning rate, optimizer, and batch size, are kept identical across all models.

Figure 6 presents the confusion matrices of different models. Note that the darker shades in Fig. 6 are only used to represent larger quantities of each category. STMNN, EMCAN, and STA-SAW clearly outperform the MKFCNN, PCNN-LSTM, SCNN-LSTM, and CNN-GRU, indicating their stronger adaptability to imbalanced data. However, the performance of these models remains uneven. For example,

the STA-SAW achieves high accuracy on Label 0 but underperforms on Label 3. In contrast, the proposed MCNN-LSTM model maintains consistently good performance across all classes, particularly in the minority class. Figure 7 presents the comparison of recall and F1-score of each class by different models. While all models perform similarly on Label 0 (due to abundant samples), significant gaps emerge for classes with limited data. The proposed MCNN-LSTM model yields the highest recall for Labels 1, 2, 3, and 5, and outperforms all models in F1-score for all labels—particularly for the three rarest Labels, i.e., Labels 2, 3, and 4, showing an excellent balance between precision and recall.

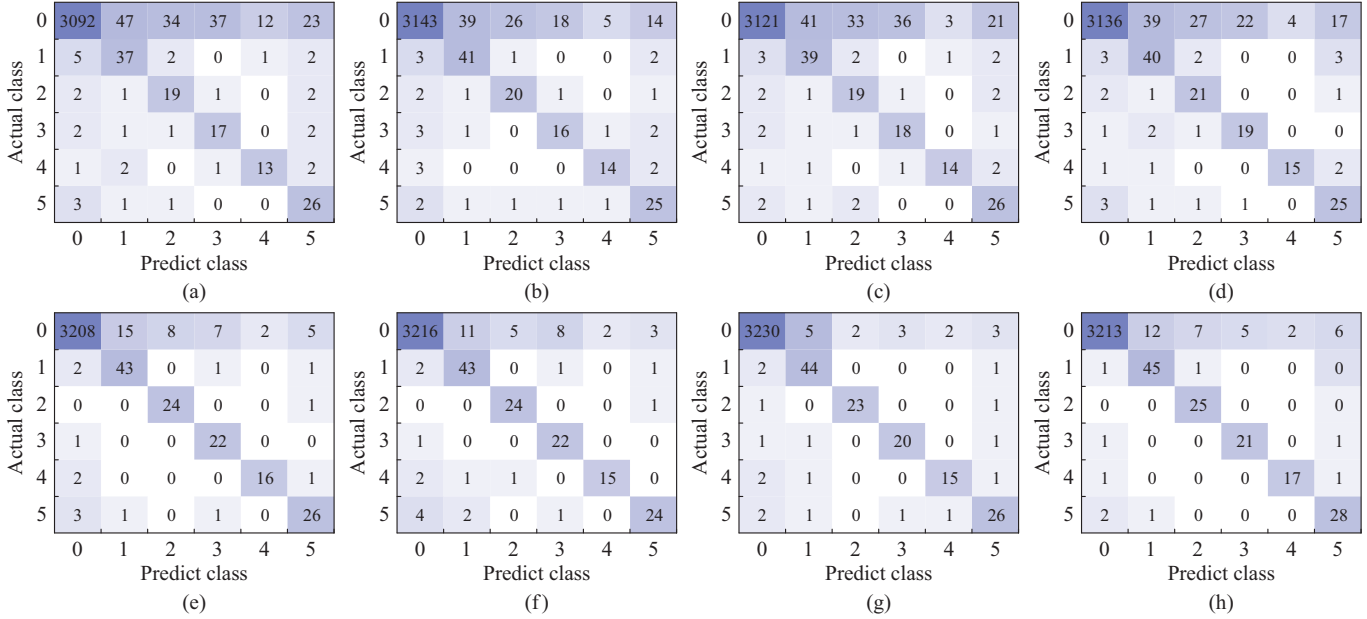


Fig. 6. Confusion matrices of different models. (a) MKFCNN. (b) PCNN-LSTM. (c) SCNN-LSTM. (d) CNN-GRU. (e) STMNN. (f) EMCAN. (g) STA-SAW. (h) MCNN-LSTM.

The comparison of diagnostic results from different models is shown in Supplementary Material A Fig. SA5. In Fig. SA5(a), a bar chart visually demonstrates the scores of each model across individual metric. The proposed MCNN-LSTM model takes the lead across multiple metrics. Specifically, the Macro-recall score of the proposed MCNN-LSTM model is 0.943, which is 2.83% higher than that of STMNN. The proposed MCNN-LSTM model also achieves the highest scores in Macro-F1 and G-mean. The radar chart in Fig. SA5(b) further confirms its consistent advantage across individual metric. Collectively, results from Figs. 6, 7, and SA5 validate that the proposed MCNN-LSTM model delivers more balanced, robust, and accurate diagnostic performance under imbalanced data compared with existing models.

E. Discussion

The proposed MCNN-LSTM model achieves robust fault diagnosis performance under imbalanced data without relying on data augmentation, distinguishing it from many existing models. This capability arises from the joint design of the input representation, model architecture, and learning objective. The IMTS explicitly preserves the spatial topology among SCADA sensors and the temporal evolution patterns

through a dual-label matrix structure, offering a more informative basis for feature learning than conventional univariate sequences. The MCNN captures cross-variable spatial correlations that serve as diagnostic priors, while the stacked LSTM network layers model temporal dynamics at multiple time scales, enabling deep spatiotemporal fusion. Furthermore, the IWCE loss, derived using KDE-based sample probabilities, adaptively assigns higher weights to minority and hard-to-classify samples during training, thereby improving recall for rare fault types without manual tuning.

Nevertheless, several limitations warrant attention. The multi-branch architecture introduces additional computational complexity compared with single-path networks, which may affect deployment in latency-sensitive scenarios. In addition, the learned cross-variable spatial correlations, though effective, lack clear physical interpretability, which limits the transparency of diagnostic decisions for domain engineers. Finally, the evaluation is based on data collected from a single wind farm, and the generalization across different WTs, manufacturers, or operation environments has not been verified. These aspects highlight important directions for future improvement.

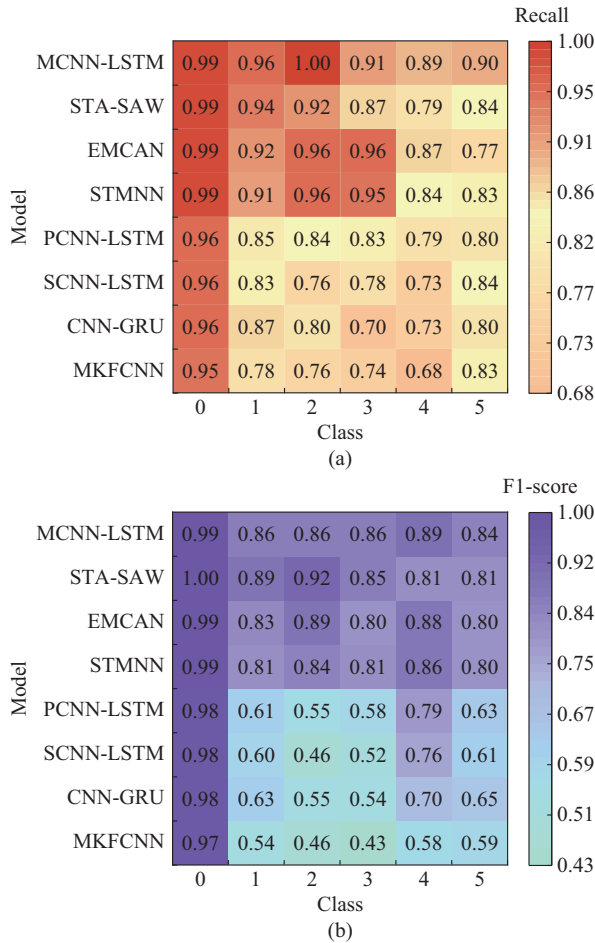


Fig. 7. Comparison of diagnostic results of various faults by different models. (a) Recall. (b) F1-score.

V. CONCLUSION

This paper proposes spatiotemporal feature extraction from IMTS via MCNN-LSTM for WT fault diagnosis, which extracts spatiotemporal features from dual-label IMTS and adopts an IWCE loss to highlight the minority class.

Experimental results on real SCADA data verify that the proposed MCNN-LSTM model achieves superior Macro-F1 and G-mean performances without synthetic data augmentation, providing a practical solution for WT intelligent maintenance. Future research will focus on improving the cross-site generalization of the proposed MCNN-LSTM model, lightweight design and diagnostic interpretability for engineering deployment.

REFERENCES

- [1] Q. Zhong, S. Liu, C. Liu *et al.*, "Fault diagnosis of wind turbine gearboxes based on multisource signal fusion," *IEEE Transactions on Instrumentation and Measurement*, vol. 74, no. 4, pp. 1-13, Apr. 2025.
- [2] L. Sheng, C. Li, M. Gao *et al.*, "A review of SCADA-based condition monitoring for wind turbines via artificial neural networks," *Neurocomputing*, vol. 633, no. 6, p. 129830, Jun. 2025.
- [3] W. Gu, Z. Chen, Q. Li *et al.*, "Torque limit-based inertial control method based on delayed support for primary frequency control of wind turbines," *Journal of Modern Power System and Clean Energy*, vol. 12, no. 2, pp. 561-570, Jun. 2023.
- [4] L. Zhang, H. Zhang, and G. Cai, "The multiclass fault diagnosis of wind turbine bearing based on multisource signal fusion and deep learning generative model," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, p. 3514212, May 2022.
- [5] Y. Zhu, C. Zhu, J. Tan *et al.*, "Fault detection of offshore wind turbine gearboxes based on deep adaptive networks via considering Spatio-temporal fusion," *Renewable Energy*, vol. 200, no. 11, pp. 1023-1036, Nov. 2022.
- [6] H. Lai, J. Chen, K. Xiahou *et al.*, "Fault diagnosis of wind turbine bearing based on parameter-optimized VMD and multi-feature fusion," *Arabian Journal for Science and Engineering*, vol. 50, no. 8, pp. 5425-5438, Apr. 2025.
- [7] T. Roy, D. K. Ray, and S. Chattopadhyay, "Skewness driven DFIG's DC link bus short and open circuit faults diagnosis using rotor and stator currents," *IEEE Transactions on Energy Conversion*, vol. 39, no. 1, pp. 107-115, Mar. 2024.
- [8] X. Zeng, M. Yang, C. Feng *et al.*, "Data-driven anomaly detection method based on similarities of multiple wind turbines," *Journal of Modern Power System and Clean Energy*, vol. 12, no. 3, pp. 803-818, May 2024.
- [9] Y. Zhu, B. Xie, A. Wang *et al.*, "Wind turbine fault detection and identification via self-attention-based dynamic graph representation learning and variable-level normalizing flow," *Reliability Engineering & System Safety*, vol. 253, no. 1, p. 110554, Jan. 2025.
- [10] A. Wang, Y. Pei, Y. Zhu *et al.*, "Wind turbine fault detection and identification through self-attention-based mechanism embedded with a multivariable query pattern," *Renewable Energy*, vol. 211, pp. 918-937, Jul. 2023.
- [11] Y. Zhu, C. Zhu, J. Tan *et al.*, "Anomaly detection and condition monitoring of wind turbine gearbox based on LSTM-FS and transfer learning," *Renewable Energy*, vol. 189, no. 4, pp. 90-103, Apr. 2022.
- [12] L. Xiang, X. Yang, A. Hu *et al.*, "Condition monitoring and anomaly detection of wind turbine based on cascaded and bidirectional deep learning networks," *Applied Energy*, vol. 305, p. 117925, Jan. 2022.
- [13] J. Liu, X. Wang, F. Xie *et al.*, "Condition monitoring of wind turbines with the implementation of spatio-temporal graph neural network," *Engineering Applications of Artificial Intelligence*, vol. 121, p. 106000, May 2023.
- [14] X. Wen, K. Yang, P. Xin *et al.*, "A novel SE-weighted multi-scale hedging CNN approach for fault diagnosis of wind turbine," *Measurement Science And Technology*, vol. 34, no. 9, p. 95905, Jun. 2023.
- [15] P. Xin, X. Zhang, C. Yuan *et al.*, "Fault diagnosis of wind turbine based on CNN-LSTM parallel network model," *Academic Journal of Engineering and Technology Science*, vol. 6, no. 12, pp. 115-121, Dec. 2023.
- [16] J. Zhan, C. Wu, C. Yang *et al.*, "Condition monitoring of wind turbines based on spatial-temporal feature aggregation networks," *Renewable Energy*, vol. 200, no. 11, pp. 751-766, Nov. 2022.
- [17] S. Sun, W. Hu, Y. Liu *et al.*, "Matching contrastive learning: an effective and intelligent method for wind turbine fault diagnosis with imbalanced SCADA data," *Expert Systems With Applications*, vol. 223, p. 119891, Aug. 2023.
- [18] H. Yi, Q. Jiang, G. Yan *et al.*, "Imbalanced classification based on minority clustering synthetic minority oversampling technique with wind turbine fault detection application," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 9, pp. 5867-5875, Sept. 2021.
- [19] M. Qian and Y. Li, "A weakly supervised learning-based oversampling framework for class-imbalanced fault diagnosis," *IEEE Transactions on Reliability*, vol. 71, no. 1, pp. 429-442, Mar. 2022.
- [20] S. Yang, Y. Zhou, X. Chen *et al.*, "Fault diagnosis of wind turbines with generative adversarial network-based oversampling method," *Measurement Science and Technology*, vol. 34, no. 4, p. 044004, Jan. 2023.
- [21] Y. Zhang, Y. Zhang, Y. Zhang *et al.*, "Multi-class data augmentation and fault diagnosis of wind turbine blades based on ISOMAP-CGAN under high-dimensional imbalanced samples," *Renewable Energy*, vol. 243, p. 122609, Apr. 2025.
- [22] X. Zeng, L. Xia, C. Feng *et al.*, "An improved conditional generative adversarial method for wind turbine gearbox fault diagnosis with imbalance data," *IEEE Sensors Journal*, vol. 24, no. 23, pp. 38727-38739, Nov. 2024.
- [23] N. Anantrasichai and D. Bull, "Defectnet: multi-class fault detection on highly-imbalanced datasets," in *Proceedings of 2019 IEEE International Conference on Image Processing*, Taipei, China, Sept. 2019, pp. 2481-2485.
- [24] S. Sun, T. Wang, and F. Chu, "A multi-learner neural network approach to wind turbine fault diagnosis with imbalanced data," *Renewable Energy*, vol. 208, no. 5, pp. 420-430, May 2023.
- [25] J. Guo, Y. Liu, Y. Li *et al.*, "Imbalanced learning for wind turbine

- blade icing detection via spatio-temporal attention model with a self-adaptive weight loss function,” *Expert Systems with Applications*, vol. 229, p. 120428, Nov. 2023.
- [26] Q. He, Y. Pang, G. Jiang *et al.*, “A spatio-temporal multiscale neural network approach for wind turbine fault diagnosis with imbalanced SCADA data,” *IEEE Transactions on Industrial Informatics*, vol. 17, no. 10, pp. 6875–6884, Oct. 2021.
- [27] M. Modirrousta, M. Shoorehdeli, M. Yari *et al.*, “Imbalanced classification in faulty turbine data: new proximal policy optimization,” *IET Collaborative Intelligent Manufacturing*, vol. 6, no. 3, p. 12144, Jan. 2023.
- [28] J. Ma, Y. Fu, T. Cheng *et al.*, “SCADA data-driven spatio-temporal graph convolutional neural network for wind turbine fault diagnosis,” *IEEE Transactions on Instrumentation and Measurement*, vol. 74, p. 3521910, Mar. 2025.
- [29] J. Guo, C. Chong, P. Abreu *et al.*, “Reparameterization convolutional neural networks for handling imbalanced datasets in solar panel fault classification,” *Engineering Applications of Artificial Intelligence*, vol. 150, p. 110541, Jun. 2025.
- [30] Z. Xu, M. Bashir, Y. Yang *et al.*, “Multisensory collaborative damage diagnosis of a 10 MW floating offshore wind turbine tendons using multi-scale convolutional neural network with attention mechanism,” *Renewable Energy*, vol. 199, no. 11, pp. 21–34, Nov. 2022.
- [31] Z. Xu, X. Mei, X. Wang *et al.*, “Fault diagnosis of wind turbine bearing using a multi-scale convolutional neural network with bidirectional long short term memory and weighted majority voting for multi-sensors,” *Renewable Energy*, vol. 182, pp. 615–626, Jan. 2022.
- [32] C. Sudre, W. Li, T. Vercauteren *et al.*, “Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations,” in *Proceedings of International Workshop on Deep Learning in Medical Image Analysis*, Québec City, Canada, Sept. 2017, pp. 240–248.
- [33] H. Majzoub, I. Elgedawy, O. Akaydin *et al.*, “HCAB-SMOTE: a hybrid clustered affinitive borderline SMOTE approach for imbalanced data binary classification,” *Arabian Journal for Science and Engineering*, vol. 45, no. 4, pp. 3205–3222, Apr. 2020.
- [34] M. Kumari and N. Subbarao, “A hybrid resampling algorithms SMOTE and ENN based deep learning models for identification of marburg virus inhibitors,” *Future Medicinal Chemistry*, vol. 14, no. 10, pp. 701–715, May 2022.
- [35] Y. Pang, G. Jiang, Q. He *et al.*, “Multi kernel fusion convolutional neural network for wind turbine fault diagnosis,” in *Proceedings of 2019 Chinese Automation Congress*, Hangzhou, China, Nov. 2019, pp. 2871–2876.
- [36] Z. Kong, B. Tang, L. Deng *et al.*, “Condition monitoring of wind turbines based on spatio-temporal fusion of SCADA data by convolutional neural networks and gated recurrent units,” *Renewable Energy*, vol. 146, pp. 760–768, Feb. 2020.
- [37] T. Huang, Q. Zhang, X. Tang *et al.*, “A novel fault diagnosis method based on CNN and LSTM and its application in fault diagnosis for complex systems,” *Artificial Intelligence Review*, vol. 55, no. 2, pp. 1289–1315, Feb. 2022.
- [38] J. He, W. Huang, Y. Liu *et al.*, “Data imbalance fault diagnosis method based on an ensemble multi-scale convolutional attention network,” *Mechanical Systems and Signal Processing*, vol. 236, p. 112934, Aug. 2025.
- Xiangjun Zeng** received the B.S. and M.S. degrees in electrical engineering from Xinjiang University, Urumqi, China, in 2014 and 2017, respectively, and the Ph.D. degree in electrical engineering from Shandong University, Jinan, China, in 2022. He is currently a Lecturer in the College of Electrical Power & New Energy, China Three Gorges University, Yichang, China. His research interests include condition-based maintenance, intelligent anomaly detection, and fault diagnosis for power equipment.
- Xiangqing Fang** received the B.Eng. degree from Quancheng College, University of Jinan, Jinan, China, in 2022. He is currently pursuing the M.Eng. degree at China Three Gorges University, Yichang, China. His research interests include deep learning-based fault diagnosis of electromechanical equipment.
- Binqiao Zhang** received the B.Sc. degree in electrical technology from Henan University, Kaifeng, China, in 2001, the M.Sc. degree in computer application from China Three Gorges University, Yichang, China, in 2007, and the Ph.D. degree in system analysis and integration from Huazhong University of Science and Technology, Wuhan, China, in 2017. Since 2001, he has been engaged in teaching and research activities at the College of Electrical Engineering and New Energy, China Three Gorges University. Currently, he serves as an Associate Professor at the same institution. His research interests include the informatization of power industry and operation simulation of hydropower station.
- Chen Feng** received the B.S. degree from Shandong University of Science and Technology, Qingdao, China, in 2012, and the M.S. degree from China University of Mining and Technology, Beijing, China, in 2016. From 2017 to 2018, he was a Visiting Scholar with the Ohio State University, Columbus, USA. He received the Ph.D. degree in electrical engineering from China University of Mining and Technology, in 2020. He is now the Associate Professor in Shandong University of Science and Technology. His research interests include anomaly detection, machine learning-based maintenance, and intelligent fault diagnosis for power equipment.
- Shengyuan Zhou** received the B.Sc. degree from the School of Automation and Electrical Engineering, University of Jinan, Jinan, China, in 2018, the M.Sc. degree from the School of Electrical Engineering, Shandong University, Jinan, China, in 2021. Currently, he is working at Jinan Power Supply Company of State Grid Shandong Electric Power Company, Jinan, China. His main research interests include condition assessment and operation and maintenance of transmission line.