



A hybrid 3DSE-CNN-2DLSTM model for compound fault detection of wind turbines

Tian Wang, Linfei Yin*

School of Electrical Engineering, Guangxi University, Nanning, Guangxi 530004, China



ARTICLE INFO

Keywords:

Wind turbine
Compound fault detection
Spatial-temporal correlation
SE attention
CNN-LSTM

ABSTRACT

In recent years, intelligent fault detection methods have achieved dramatic results for wind power generation. However, a majority of the available intelligent detected methods can only detect single faults. The supervisory control and data acquisition (SCADA) system is widely applied in wind turbines (WTs). SCADA data is a time-stream of variable data. Considering spatial correlation and temporal correlation among SCADA data, this study proposes a WT compound fault detection model combining a three-dimensional squeeze-excitation convolutional neural network (3DSE-CNN), and a two-dimensional long and short-term memory network (2DLSTM). The proposed 3DSE-CNN-2DLSTM model consists of three parts: data preprocessing, spatiotemporal feature extraction, and fault detection. First, one-dimensional data is converted into a 2D image by a sliding window method. Then, in the proposed 3DSE-CNN model, the SE attention module is applied to enhance the convolutional channel information and extract spatial features. The proposed 2DLSTM model extracts spatiotemporal fusion features. Finally, the label of the fault category is obtained by argmax function. To validate 3DSE-CNN-2DLSTM, a 3 MW WT SCADA dataset from the south coast of Ireland is applied in the experiment. The proposed 3DSE-CNN-2DLSTM model achieves specific compound fault detection and improves detection accuracy.

1. Introduction

With global carbon neutrality targets being advanced, wind energy is gaining increasing attention as a renewable energy source (Ritchie et al., 2022). The global installed capacity of wind turbines (WTs) is growing year by year. Statistics for 2022 show that the world has attained 824 GW of total installed wind power capacity (Ritchie et al., 2022). Especially, China has the largest installed capacity of WTs (Ritchie et al., 2022). The harsh operating environment (Jung et al., 2022) and complex operating systems of the WTs (Pryor and Barthelmie, 2021) have resulted in frequent faults. WTs faults have now been established as a topical issue for academic research (Yang et al., 2022). The researcher proposed intellectual fault detection methods, which are vital for ensuring the safe and effective operation of WTs (Subbulakshmi et al., 2022).

The relevant sensors logged by supervisory control and data acquisition (SCADA) can indicate the operational status of WTs (Aziz et al., 2021). A spatial correlation exists between SCADA data variables. For

example, SCADA data containing fault attribute information has been applied to the resnet50 network, enabling WT power converter fault detection (Xiao et al., 2021). The correlations between SCADA data variables were quantified and extracted and then utilized for WT subsystem fault detection studies via the radar plot method, with generator faults, converter faults, and pitch system faults (Liu et al., 2020). SCADA data without historical data information has been employed for fault isolation (Liu et al., 2022). Thus, by mining the spatial correlation between SCADA data variables, WT fault detection has been achieved. WT SCADA status variables associated with specific subsystems or components (e.g., active power, temperature, and pressure) typically went through a sub-healthy period of anomalies or fault accumulation. These SCADA status variables in the sub-health period had temporal correlations. For example, SCADA data is converted into temporally correlated two-dimensional (2D) image data by preprocessing techniques. The 2D image data has been applied in deep learning (DL) structures for small anomaly classification of WTs (Wang et al., 2023). SCADA data has been applied to the gated recirculation unit (GRU) neural network for fault

Abbreviations: DL, Deep learning; FL, Focal loss; GRU, Gate recurrent units; 1D, One-dimensional; RTU, Remote terminal unit; SCADA, Supervisory control and data acquisition; SVM, Support vector machines; TCN, Temporal convolutional network; 3DSE-CNN, Three-dimensional squeeze-excitation CNN; 2DLSTM, Two-dimensional LSTM; WEC, Wind energy converter; WT, Wind turbine.

* Corresponding author.

E-mail address: yinlinfei@gxu.edu.cn (L. Yin).

<https://doi.org/10.1016/j.eswa.2023.122776>

Received 19 April 2023; Received in revised form 28 October 2023; Accepted 26 November 2023

Available online 30 November 2023

0957-4174/© 2023 Elsevier Ltd. All rights reserved.

detection in the WT main bearings (Encalada-Davila et al., 2022). SCADA data has been employed in non-parametric statistical testing methods for the operation and detection of WTs (Dao, 2022). The above WT fault detection methods focused on the spatial structure or temporal order of SCADA data. However, most existing research methods did not capture the spatiotemporal correlation of SCADA data properly (Aha-konye et al., 2023) and did not fully exploit potential information in SCADA data (He et al., 2021). Existing research methods limited the work on WT fault detection.

To ensure the reliable operation of WTs, several fault detection algorithms have been applied to WTs (Lei et al., 2020). Traditional machine learning focused on manual feature extraction and classification by classifiers. Traditional machine learning relied on expertise and high feature requirements, e.g., extreme learning machines (Li et al., 2021), support vector machines (SVM) (Dhiman et al., 2021), and decision trees (Yang et al., 2021). DL focuses on automatic feature learning and reduces the reliance on expert experience, e.g., stacked autoencoder (Wang et al., 2022), convolutional neural networks (CNN) (Zhang et al., 2022), residual networks (Liu et al., 2023), and migration learning (Chen et al., 2021). Existing methods for smart failure detection are directed at single faults of WTs. In actual industry, WTs inevitably suffered compound faults. Compound faults have been primarily caused by the coupled and concurrent nature of WT systems. Compound faults referred to the presence of two or more single fault conditions at the same time (Yan et al., 2021). Compound fault detection algorithms mainly consist of signal processing (Deng et al., 2022) and data-driven (Xing et al., 2022). The signal processing algorithms relying on expert experience (Liao et al., 2021) have been employed to distinguish between different single faults in a compound fault. For example, a hysteresis information spectrum method has been applied to reduce signal noise and separate compound fault features (Alavi et al., 2022). A speed-free generalized demodulation method has been employed to separate the mapped frequencies for locating faults (Liu et al., 2023). The data-driven-based DL has been considered cost-effective and efficient for compound fault diagnosis (Yang et al., 2023). Data-driven algorithms generally mine compound fault information through data dimensionality reduction (Wang et al., 2021). A multi-stream self-fusion algorithm has been applied to mine the image model for heterogeneous information (Zhang et al., 2022). Most existing fault detection methods do not consider spatial-temporal correlation for high accuracy.

To derive features for spatiotemporal fusion, hybrid networks have been proposed. Hybrid networks have been proposed for CNNs and recurrent neural networks fusion (Kong et al., 2023), spatiotemporal fusion networks (STFN) (He et al., 2021), and CNN and long short-term memory (LSTM) fusion (Mustafa et al., 2023). In addition, a hybrid method for LSTM and Kullback-Leibler scattering has been applied to monitor variables and capture temporal information for WTs state monitoring (Wu and Ma, 2022). Feature engineering and MuSNet have been employed to extract highly correlated variables and sequence information (Zhang et al., 2022). Infrared thermography and DL fusion models have been developed to extract deep features (Attallah et al., 2023). Recently, CNN-LSTM networks have been developed for various problems including fault detection (Xiang et al., 2022), etiology identification (Ward et al., 2022), time series prediction (Kortli et al., 2022), and image classification (Wang et al., 2023). CNN-LSTM network superiority: (i) the CNN networks paid attention to the most obvious features; LSTM networks expanded their properties in temporal order; (ii) the CNN-LSTM networks extracted spatiotemporal information. The CNN-LSTM networks have been applied to solve problems including inadequate extrapolation for mechanical system fault diagnosis (Ward et al., 2022) and the automatic extraction of bearing data features (Chen et al., 2021). Therefore, the CNN-LSTM networks can effectively extract spatiotemporal information and improve fault classification performances.

Inspired by pioneering work, this study aims to design a three-dimensional squeeze-excitation (3DSE)-CNN-2DLSTM model. The

proposed 3DSE-CNN-2DLSTM model can mine deep spatiotemporal fusion features and achieve composite failure detection. The proposed 3DSE-CNN-2DLSTM model features: (i) the squeeze-excitation (SE) attention is set in the proposed 3DSE-CNN model to capture dynamical correlations of spatial variables adaptively. (ii) The spatial feature extraction among variables in SCADA data relies on the 3DSE-CNN model. The global feature extraction for each variable relies on the 2DLSTM model. (iii) The proposed 3DSE-CNN and 2DLSTM models are connected by a transformation layer, which ensures that the temporal sequence of image data transfer is not corrupted. (iv) To improve the proposed 3DSE-CNN-2DLSTM model performance reduction caused by data imbalance, a focal loss (FL) function trains the proposed 3DSE-CNN-2DLSTM model. Contributions to this work are summarized below.

- (1) WTs SCADA data preprocessing. In the methods of (Li et al., 2021; Wu and Ma, 2022; Zhu and Song, 2022; Li et al., 2019; Zhan et al., 2022), dealing with SCADA data, the multivariate time series data was converted to one-dimensional (1D) vectors, which could retain spatially correlated information. In contrast, the proposed data preprocessing method, namely the sliding window technique and the transformation technique, preserves the spatial and temporal information of the SCADA data.
- (2) The proposed 3DSE-CNN-2DLSTM model. In the CNN-LSTM networks described in (Wu and Ma, 2022; Ward et al., 2022; Kortli et al., 2022; Wang et al., 2023; Li et al., 2019) CNN output layers connect flatten layers. The flattened layers flatten the acquired features into a 1D vector. Although 1D vectors have the implicit features of time series, some spatial features are lost. In this study, the 3D image feature data outputs from the 3DSE-CNN model and the superimposed 2D original image in the proposed 3DSE-CNN-2D LSTM model are passed through a transformation layer to obtain the 2D image data. The obtained 2D image data retains the spatiotemporal fusion information.
- (3) The specific compound fault detection of WTs is implemented. The source datasets selected for this study have been applied (He et al., 2021; Zhu and Song, 2022; Li et al., 2019; Zhu and Zhang, 2021; Leahy et al., 2016; Leahy et al., 2018), where a single fault detection of WTs was implemented in the models. However, the models did not consider the compound faults of WTs. In this study, a 3DSE-CNN-2DLSTM model for some specific compound fault detection in WTs is proposed.
- (4) To the best of the knowledge of the author, the proposed 3DSE-CNN and 2DLSTM in this study are utilized for the first time in this dataset of WTs from a biopharmaceutical plant near the Irish coast. The 3DSE-CNN-2DLSTM is compared with CNN, LSTM, GRU, and temporal convolutional network (TCN) models for experiments. The 3DSE-CNN-2DLSTM outperforms other comparative models.

This study consists of the following main sections. Section 2 is mainly concerned with the description and labeling of SCADA data. Section 3 introduces the theory of the 3DSE-CNN model, 2DLSTM model, and data preprocessing, and establishes the 3DSE-CNN-2DLSTM model. Section 4 analyses and discusses the 3DSE-CNN-2DLSTM model for composite fault detection. This work is abstractly summarized in Section 5.

2. Data

The dataset utilized for this research is derived from a 3 GW WT SCADA system in Ireland. The dataset has been utilized extensively in studies (He et al., 2021; Zhu and Song, 2022; Zhu and Zhang, 2021; Leahy et al., 2016; Leahy et al., 2018).

2.1. Data description

The SCADA system typically utilizes vibration sensors to record mechanical vibrations, and electrical parameter monitoring equipment to record information about the electrical system. The vibration sensor is

Table 1
Partial sample data.

Date	Time	Ava windspeed m/s	Ava power kW	Spinner temp °C	Front bearing temp °C	Nacelle temp °C
31/07/2014	18:50:00	7.9	913	25	29	24
31/07/2014	19:00:00	8.6	1376	25	29	25
31/07/2014	19:10:00	7.4	900	25	29	24
31/07/2014	19:20:00	8.1	1140	24	29	24

Table 2
Partial status information.

Date	Time	Status	Text
24/04/2014	12:37:38	0:0	Turbine in operation
14/05/2014	14:41:31	9:3	Generator heating: hygrostat inverter
14/05/2014	19:49:23	2:1	Lack of wind
04/06/2014	08:17:53	60:11	Mains failure: undervoltage L1
05/06/2014	17:33:16	62:505	Feeding fault
09/06/2014	00:00:00	228:100	Timeout warn message

Table 3
SCADA Data variables.

No	Variable name	No	Variable name
A ₁	Average wind speed	A ₁₂	Nacelle ambient temp
A ₂	Average rotation	A ₁₃	Nacelle temp
A ₃	Average power	A ₁₄	Nacelle cabinet temp
A ₄	Average reactive power	A ₁₅	Main carrier temp
A ₅	Average blade angle	A ₁₆	Rectifier cabinet temp
A ₆	Spinner temp	A ₁₇	Ambient temp
A ₇	Front bearing temp	A ₁₈	Tower temp
A ₈	Rear bearing temp	A ₁₉	Control cabinet temp
A ₉	Pitch cabinet blade temp	A ₂₀	Transformer temp
A ₁₀	Rotor temp	A ₂₁	Inverter average cabinet temp
A ₁₁	Stator temp	A ₂₂	Inverter std dev cabinet temp

installed on mechanical components for health monitoring of mechanical systems such as generators, gearboxes, and bearings. The electrical parameter monitoring equipment is utilized to monitor the performance and stability of electrical systems, mainly current sensors, voltage sensors, power factor monitoring equipment, and temperature sensors. These devices realize real-time monitoring and data collection of the operating status of WTs through the SCADA system. The SCADA system in this study records operational data for WTs between 2014 and 2015 (He et al., 2021). The SCADA data is divided into three different datasets: operational, status, and alarm datasets.

The operational dataset records 40,927 pieces of data. Each piece of data has a timestamp. The data is logged with 10-minute intimacy. Each operation consists of 64 variables. The variables include mainly: wind power variables, component temperatures, electrical measurements, and operating conditions. The operational data reflects the state and conditions of WT during operation. Some of the multivariate time series sample data (Table 1) mainly include values for variables such as time stamp, wind speed, power, and temperature. The operational data is original data and may exist with some noise or missing values. The original data is pre-processed and cleaned for subsequent analysis and modeling.

The status dataset contains the operational status data of the WT, divided into wind energy converter (WEC) and remote terminal unit (RTU) data. When either the WEC or RTU status changes, a status message has an associated “primary status: secondary status” code. Several status messages (Table 2) consist mainly of date, timestamp, status code, and status description.

The warning dataset records general information about the WT. Warning messages are usually not directly related to the operational safety of the WT. Warning messages are mainly utilized to monitor the status and performance of the WT. Warning signals, in a similar way to the status database, capture different types of news, e.g., WT speed,

ambient temperature, and wind direction. The warning messages are recorded in the same way as status messages. Although the warning messages are usually not directly related to the operational safety of the WT, a record of the warning messages is important to ensure the long-term stability of the WT.

According to correlation analysis (Zhang et al., 2021), a total of 22 variables (Table 3) are selected for this study and normalized for each variable after removing invalid data. At different stages of the WT, each variable shows different fluctuation trends (Fig. 1) indicating that the trend of each variable does not directly determine WT status. To summarize, predicting the operational status of WT requires a combination of several variables, which are analyzed and processed with complex algorithms and models.

2.2. Data labeling

To train a classifier, correct labeling of the data is very significant. In this study, the SCADA operational dataset is annotated in combination with the status dataset and the warning dataset. This study counts the eight status codes that occur with high frequency in the SCADA system. State code 62 is a feeding fault, which usually refers to a problem with the feeding cable of the WTs. The feeding fault may include problems such as broken cables, poor connections, and damaged electrical connectors. The feeding fault affects the transmission of electrical energy and causes the generator to fail to operate properly. Status code 80 is an excitation fault, which usually refers to a problem with the generator excitation system. The excitation fault may be caused by damage to the excitation winding, faults in the regulation system, and abnormal excitation current. The occurrence of excitation faults affects the voltage stability of the generator and may lead to power system instability. Status code 228 is an air-cooling fault, which usually refers to problems related to the air circulation and internal temperature circulation system. Common causes of air-cooling faults are clogged air circulation passages, fan failures, and coolant problems. Status code 60 is a mains fault, which usually means that there is a problem with the mains power supply to the WTs. The mains fault may be caused by a faulty cable, a faulty switch, or a faulty other electrical component. Status code 9 is a generator heat fault, usually a generator overheating problem. A compound fault is two or more faults occurring at the same time. The two compound faults counted in this study are the feeding and excitation fault and, the mains and air-cooling fault. Status code 62–80 is feeding and excitation fault. The simultaneous occurrence of both a feeding fault and excitation fault in a wind turbine usually means that there is a serious problem with the electrical system of the generator. While these two types of faults are usually independent, in some cases they may interact with each other, especially under extreme conditions or due to system design or manufacturing defects. Status code 60–228 is a mains fault and an air-cooling fault. The simultaneous occurrence of a mains fault and an air-cooling fault in WTs is usually the result of a mains fault triggering other faults in the control system or the electrical system, which may cause the air-cooling system to fail to operate properly, and thus trigger an air-cooling fault. Table 4 summarizes the data information for each health condition. Except for the single fault category, and compound fault category, no fault belongs to the normal category.

In the SCADA system, the wind power scatter plot records the performance data of the WT. The wind power scatter plot is an essential indicator for assessing the operational status of the WT. The

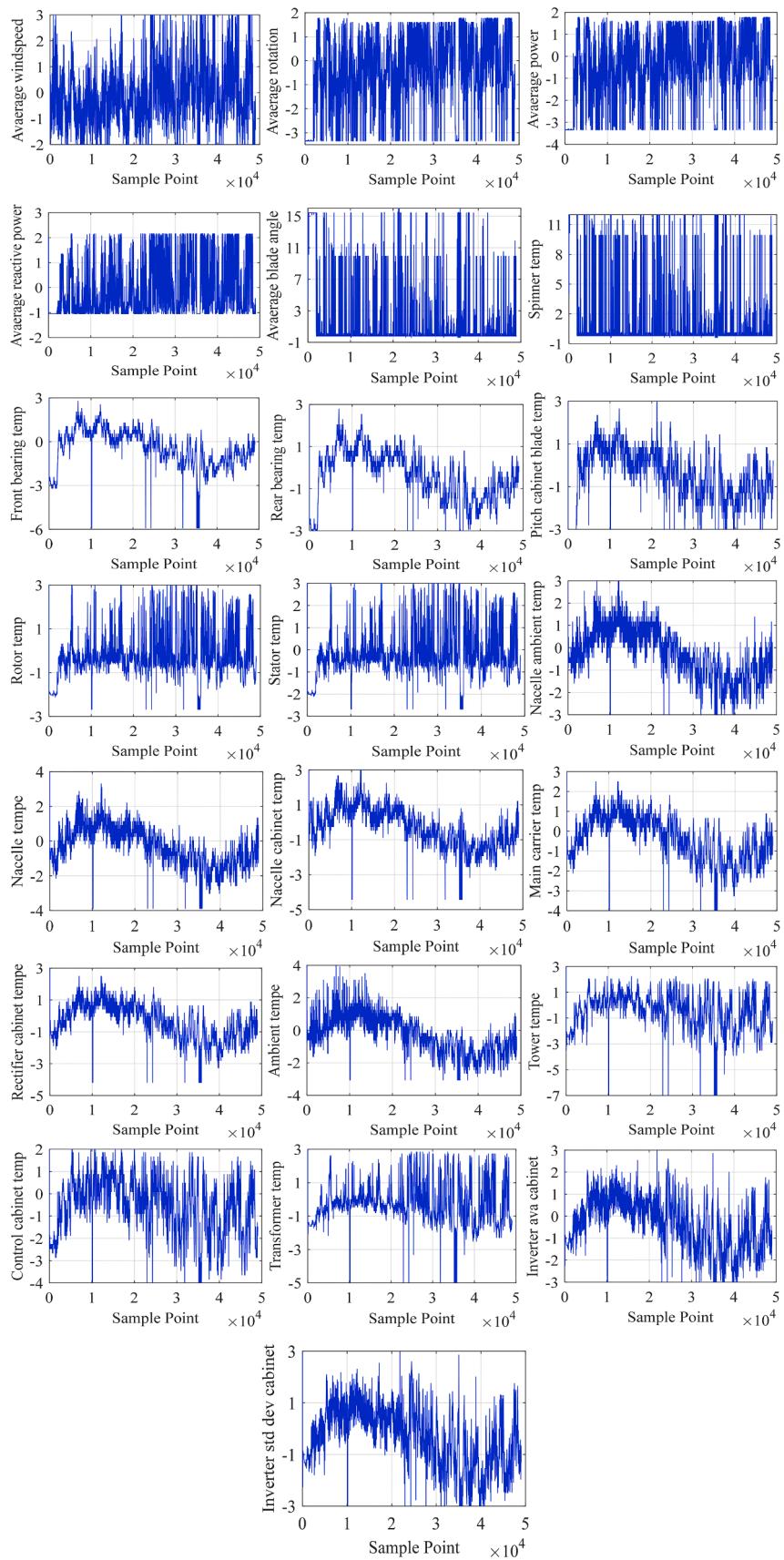


Fig. 1. Fluctuating trends after normalization of SCADA data variables.

Table 4
Fault data statistics.

Fault	State code	Total number of data	Fault statement	Label
No fault	0	32,049	Turbine in operation	[0,0,0,0,0,1]
Feeding fault	62	162	No zero-crossing inverter	[1,0,0,0,0,0]
Excitation fault	80	77	Overshoot dc-link	[0,0,0,1,0,0]
Air-cooling fault	228	50	Malfunction air-cooling	[0,0,1,0,0,0]
Mains fault	60	8	Undervoltage	[0,1,0,0,0,0]
Generator heating fault	9	44	Hygrostat inverter	[0,0,0,1,0,0]
Feeding and excitation fault	62–80	101	Two faults occur simultaneously	[1,0,0,1,0,0]
Mains and air-cooling fault	60–228	12	Two faults occur simultaneously	[0,1,1,0,0,0]

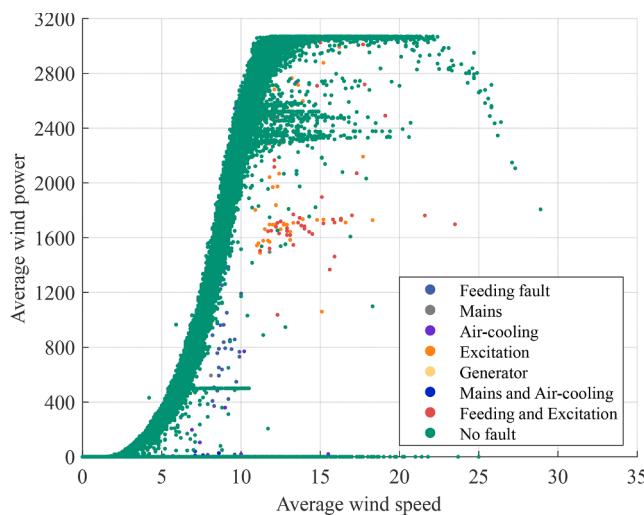


Fig. 2. WT wind power scatter diagram.

visualization of the average power and average wind speed for the eight different status messages (Fig. 2) shows that the different fault status messages are mixed with normal messages. The wind power diagram does not allow direct determination of fault conditions. As a result, system operating data, warning messages, and status data recorded by SCADA alone cannot accurately identify faults, especially compound fault categories. To achieve compound fault detection in WTs, this study is based on supervised learning to train and test the model, the training set is the dataset utilized to train the model. The training dataset contains the input data and the corresponding labels. The test set also contains input data and corresponding labels, but these labels are not utilized during model training. The test set compares the predicted results with the true labels of the test set to compute the various performance metrics of the model. This study is based on supervised learning and machine learning techniques to learn fault patterns from real SCADA data. The objective is to extract fault features to improve the accuracy and efficiency of identifying various fault categories.

3. Proposed 3DSE-CNN-2DLSTM model

The proposed 3DSE-CNN-2DLSTM model aims to learn the space-time fusion characteristics in SCADA data. A basic structure for the proposed 3DSE-CNN-2DLSTM model (Fig. 3) is the input layer, 3DSE-CNN model, transformation layer, 2DLSTM model, fully connected layer, and output. This proposed 3DSECNN-2DLSTM can be described

below.

3.1. Data preprocessing

Before constructing the proposed 3DSE-CNN-2DLSTM model, the SCADA data is ensured to be transformed into an appropriate form for learning spatiotemporal information. Consequently, this study utilizes a sliding window method to ensure that the input to the proposed 3DSE-CNN model is 2D original image data. At the connection between the proposed 3DSE-CNN model and the proposed 2D LSTM model, this study sets up a transformation layer to flatten 3D image data into 2D image data with a single time step. Following the transformation of all the data into 2D image data with a single time step, the proposed 2D LSTM model extracts the spatiotemporal fusion features.

3.1.1. Sliding window method

The status of the SCADA system may be related to the previous status. The SCADA data contains only spatial feature information. To learn the temporal information, the sliding window method transforms the SCADA data into 2D original image data. Fig. 4 shows the sliding window operation.

The SCADA data is $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_M]$, $\mathbf{z}_m = \{\mathbf{z}_{1m}, \mathbf{z}_{2m}, \dots, \mathbf{z}_{nm}\}$, ($\mathbf{Z} \in R^{M \times N}$, $\mathbf{z}_m \in R^N$, $\mathbf{z}_{1m} \in R$), where \mathbf{z}_m denotes a stateful vector for time m . n denotes the feature dimensions. The sliding window divides SCADA into a fixed matrix of windows along the direction of a time axis. The sliding window length is T ($0 < T \leq M$). A sliding step is β (rounding to whole numbers). The 2D original image dataset after the sliding window processing is $\mathbf{D} = \{(\mathbf{X}, \mathbf{Y})\}$, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_G]$, $\mathbf{x}_g = \{x_{1g}, x_{2g}, \dots, x_{Ng}\}$, ($\mathbf{X} \in R^{T \times N \times G}$, $\mathbf{x}_g \in R^{T \times N}$, $x_n \in R^N$), $\mathbf{Y} = \{y_1, y_2, \dots, y_G\}$ ($\mathbf{Y} \in R^G$, $y_g \in R$), \mathbf{X} denotes 2D original image data, \mathbf{Y} denotes a collection of labels. A window matrix label is the label of a vector corresponding to the last row in a matrix. At the time m_0 , the 2D subseries image data after the sliding window processing is $\mathbf{x}_{m_0} = [m_0 - w + 1 : m_0]$, labeled as $y_{m_0} = [m_0]$. The 2D original image data is utilized as the input data for the subsequent proposed 3DSE-CNN-2DLSTM model.

In this study, the window matrix is employed to capture spatial and temporal messages. The physical significance of the window matrix is to retain historical status information. In addition, the size of the window matrix is adjustable. In particular, the horizontal feature dimension n adjusts the size of the spatial information. The vertical sliding window width T adjusts the size of the temporal information. Therefore, this study controls the range of feature extraction to a certain degree by adjusting the size of the window matrix to suit the needs of different problems.

3.1.2. Transformation layer between proposed 3DSE-CNN model and proposed 2DLSTM model

The 3DSE-CNN outputs 3D image feature data. The 2DLSTM requires 2D image data as input. To transform the 3D image feature data $\tilde{\mathbf{X}}$ and the 2D original image data \mathbf{X} into a single-time step 2D image data, the transformation layer is designed in this study. Specifically, the 2D original image data \mathbf{X} and the 3D image feature data $\tilde{\mathbf{X}}$ are arranged sequentially along the feature axis in a single time step that remains constant. The input data of the proposed 2DLSTM model is represented as $\mathbf{X}_{2D} \in R^{W^* \times H_{conv}}$, $W^* = W_{conv}^j \times K$. The transformation layer operation is given in Fig. 5.

With the proposed transformation layer, this study can organically combine the proposed 3DSE-CNN and 2DLSTM models to achieve an effective fusion of spatial features and temporal information. Meanwhile, this study can control the extent to the model which can extract spatial features and temporal information by adjusting the window size. The proposed transformation layer provides greater flexibility for the model to fulfill different task requirements. Meanwhile, this study can control the extent model that allows for the extraction of spatial features and temporal information by adjusting the window size. The proposed

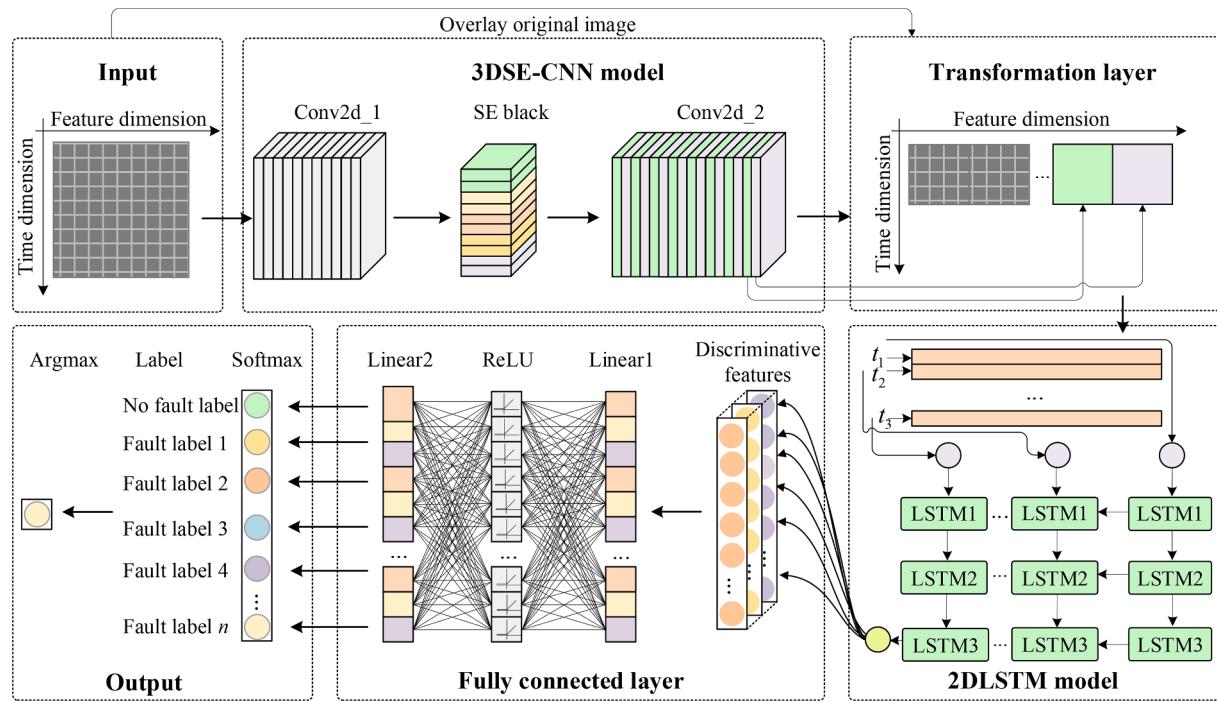


Fig. 3. Framework of the proposed 3DSE-CNN-2DLSTM model.

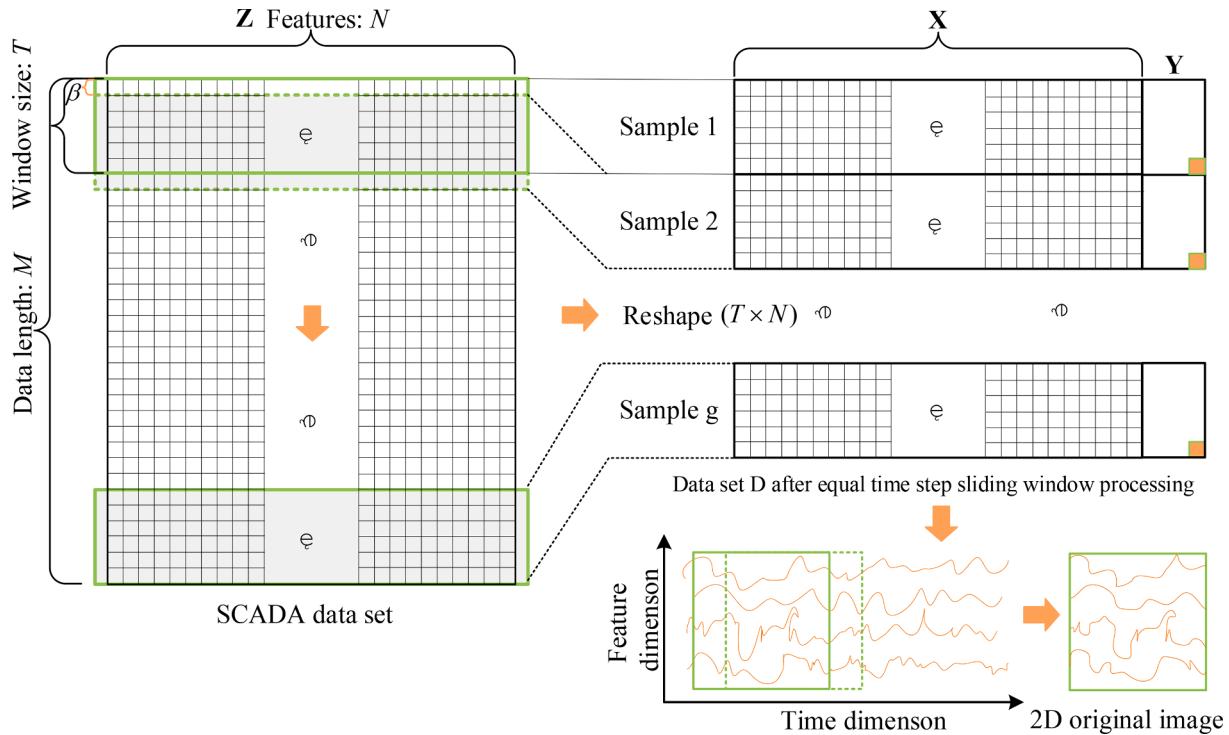


Fig. 4. Sliding window operation.

transformation layer provides greater flexibility for the model to fulfill different task requirements.

3.2. Theory overview

The proposed 3DSE-CNN-2DLSTM model mainly consists of the proposed 3DSE-CNN and 2DLSTM models. The theory of the 3DSE-CNN-2DLSTM model is given below.

3.2.1. Proposed 3DSE-CNN model

Ordinary convolution had some limitations in modeling the characterization of information channel relationships. To screen out significant information channels, this study considers the introduction of SE attention (Shu et al., 2022). The SE attention is specifically designed to enhance the correlation between the channels. The SE attention utilizes the global information of each channel to selectively enhance the favorable feature channels and suppress the unnecessary ones.

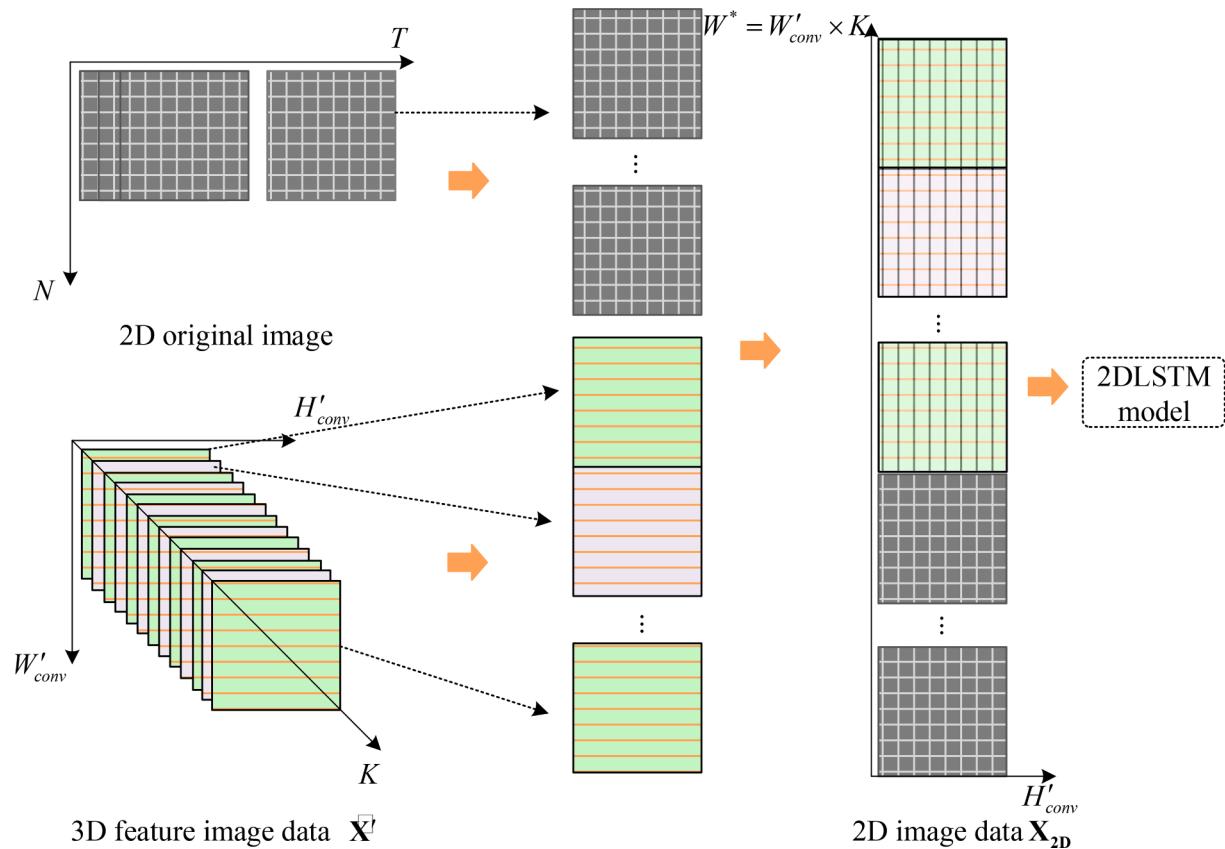


Fig. 5. Transformation layer operation.

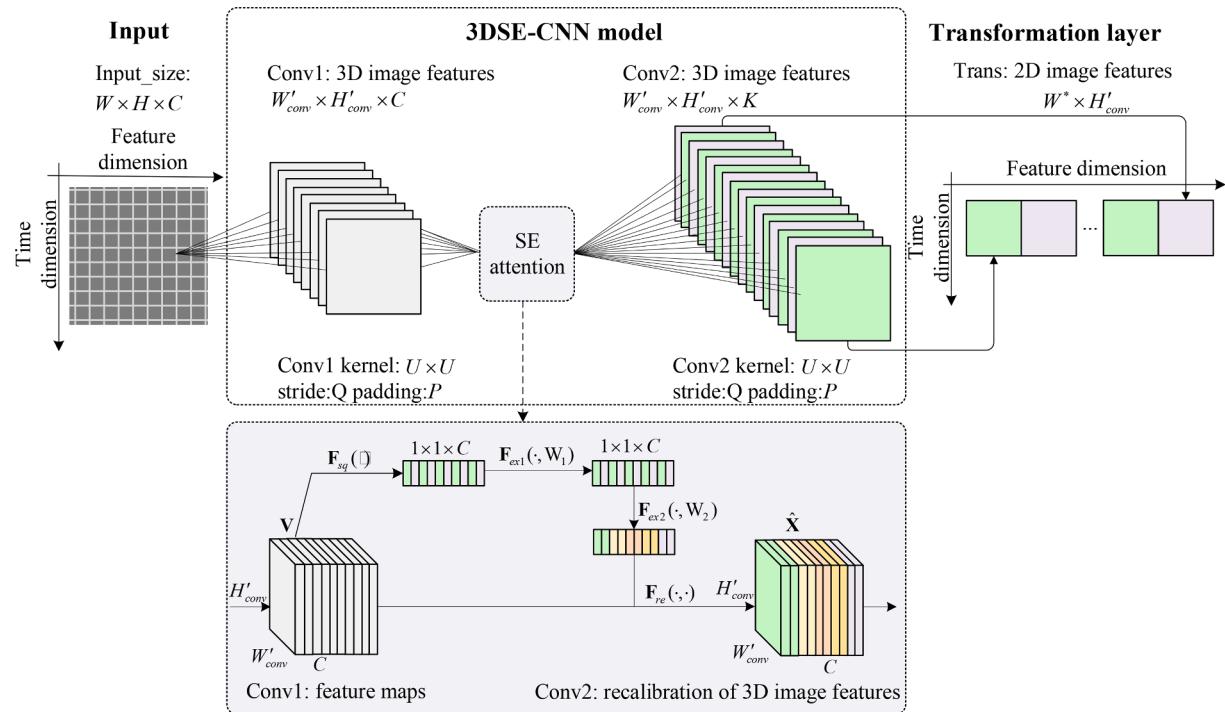


Fig. 6. Proposed 3DSE-CNN model.

Therefore, the SE attention achieves adaptive channel selection. A global average pooling captures global information. Hence, SE attention is employed as a key component in this study for channel selection and

extracting deep features.

The proposed 3DSE-CNN model in this study features local connectivity, weight sharing, and hierarchical representation. The proposed

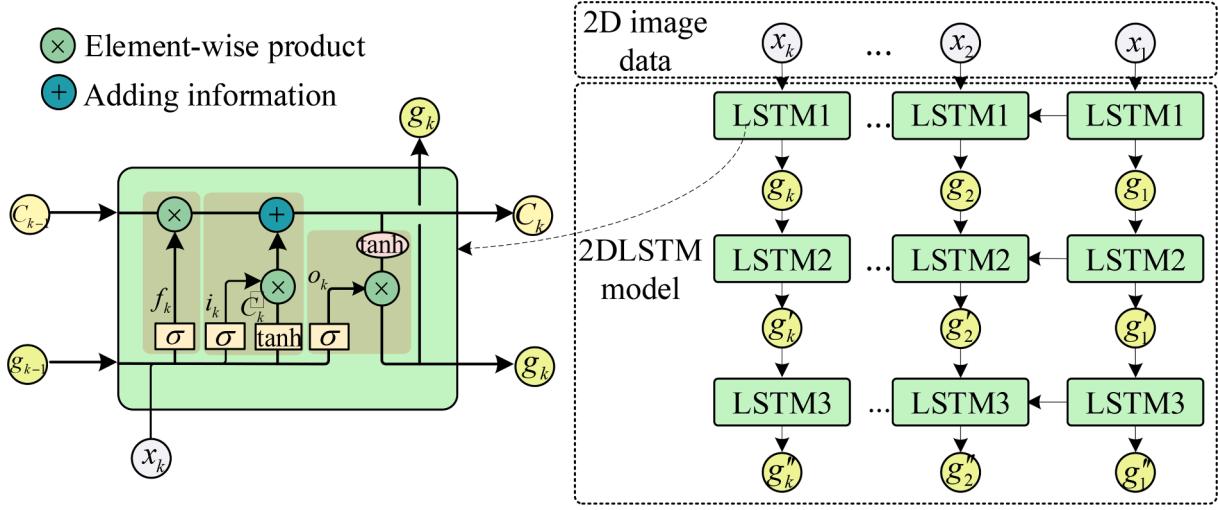


Fig. 7. Proposed 2DLSTM model.

3DSE-CNN model is utilized as the dominant network for extracting spatial features. The proposed 3DSE-CNN model (Fig. 6) is built by stacking two convolutions and one SE attention. The 3DSE-CNN model input data is $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_C]$, $\mathbf{X} \in \mathbb{R}^{W \times H \times C}$, W , H and C denotes width, height, and channels for data, separately. The 3DSE-CNN inputs are 2D image data when $C = 1$ in the first layer of convolution. The 2D original image data has a feature dimension and a time dimension on both axes.

To clearly illustrate the convolution of the proposed 3DSE-CNN model, five hyperparameters are designed, namely input channels I , output channels K , convolution kernel scale U , a convolution step Q , and the width of the image feature fill P . Before the first convolution operation, the 2D original image data $\mathbf{X} \in \mathbb{R}^{W \times H \times 1}$ is filled to obtain the 2D original image data $\mathbf{X}' \in \mathbb{R}^{(W+P) \times (H+P) \times 1}$. The convolution operation of the proposed 3DSE-CNN model is calculated as follows,

$$W'_{conv} = \frac{W - U + 2P}{Q} + 1 \quad (1)$$

$$H'_{conv} = \frac{H - U + 2P}{Q} + 1 \quad (2)$$

where W'_{conv} , H'_{conv} denote width and height after convolution, respectively.

The convolution kernel of the proposed 3DSE-CNN model is $Kernel \in \mathbb{R}^{U \times U \times C}$. The outputs of the first layer of convolution are 3D image feature data $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_C]$, $\mathbf{V} \in \mathbb{R}^{W'_{conv} \times H'_{conv} \times C}$, $\mathbf{v}_c \in \mathbb{R}^{W'_{conv} \times H'_{conv}}$. Each convolution kernel outputs one channel and the outputs are \mathbf{v}_c ; $v_{n,m}$ denote the element of \mathbf{v}_c and are calculated as follows,

$$v_{n,m} = \sum_{r=1}^U \sum_{s=1}^U \sum_{t=1}^C (Kernel_{r,s,t} \times x_{n,m}^{U \times U}) \quad (3)$$

where $x_{n,m}^{U \times U}$ denotes an element in \mathbf{X} .

To achieve adaptive channel selection within the second layer for convolution, 3DSE-CNN introduces SE attention, which consists of three basic operations: squeeze operation, excitation operation, and weight operation. The SE attention is shown in Fig. 6. The squeeze operation compresses the channel dimensions of the input tensor to obtain a summary of all channel global information. The excitation operation calculates the weights of each channel. The reweight operation performs adaptive channel selection by reweighting the original feature tensor with each channel weight. Consequently, SE attention can enhance the characterization through channel selection.

The first convolution layer output in this study is 3D image feature data \mathbf{V} . The 3D image feature data is utilized as input data for the

squeeze operation in SE attention. The squeeze operation compresses the input 3D image feature data into a real number by global pooling. The number of channels for the 3D image feature data remains the same. Thus, the image data after squeeze is transformed into $y \in \mathbb{R}^C$ y denotes the c th element, and is calculated as follows,

$$y_c = \mathbf{F}_{sq}(\mathbf{v}_c) = \frac{1}{W'_{conv} \times H'_{conv}} \sum_{r=1}^{H'_{conv}} \sum_{s=1}^{W'_{conv}} v_c(r, s) \quad (4)$$

where \mathbf{F}_{sq} is the global pooling function; $v_c(r, s)$ is the element of \mathbf{v}_c .

Secondly, the excitation operation is applied in this study to simulate the correlation of the channels. The excitation operation generates weights for each feature channel by parameters. After the excitation operation y' is calculated as follows,

$$y' = \mathbf{F}_{ex2}(\mathbf{F}_{ex1}(y, W)) = \epsilon(W_2 \mu(W_1 y)) \quad (5)$$

where μ , ϵ are ReLU function; the parameters for the descending and ascending layers are W_1 and W_2 , respectively; the ratio of both descending and ascending dimensions is l , $W_1 \in \mathbb{R}^{C \times C}$, $W_2 \in \mathbb{R}^{C \times l}$.

Lastly, with the reweight operation, this study recalibrates the 3D image feature data \mathbf{V} from the first layer of convolution in the channel dimension. The activation y' is applied in rescaling \mathbf{V} to obtain the outputs $\hat{\mathbf{x}}_c$. The formula $\hat{\mathbf{x}}_c$ is as follows,

$$\hat{\mathbf{x}}_c = \mathbf{F}_{re}(\mathbf{v}_c, y'_c) = y'_c \mathbf{v}_c \quad (6)$$

where $\hat{\mathbf{x}} = [\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \dots, \hat{\mathbf{x}}_C]$; $\mathbf{F}_{re}(\mathbf{v}_c, y'_c)$ is the weighting factor y'_c multiplied by the 3D image data \mathbf{v}_c .

In this study, the second layer of convolution repeats the first layer of convolution with K convolution kernels for a fresh set of 3D image features $\mathbf{X}' \in \mathbb{R}^{W'_{conv} \times H'_{conv} \times K}$. Finally, the 3D image feature data $\tilde{\mathbf{X}}$ is stitched together for processing.

3.2.2. Proposed 2D LSTM model

The proposed 2DLSTM model is capable to learn time-series information and extract global features. The proposed 2DLSTM model preserves the pattern of temporal changes through cell states, which enables the learning of long-term information. Fig. 7 indicates the architecture for 2DLSTM. The output subseries image feature data from 3DSE-CNN and original image data are fed into 2DLSTM for global learning. The 2DLSTM model is capable of mining deep spatiotemporal fusion features. The updated formula for the proposed 2DLSTM model is shown below.

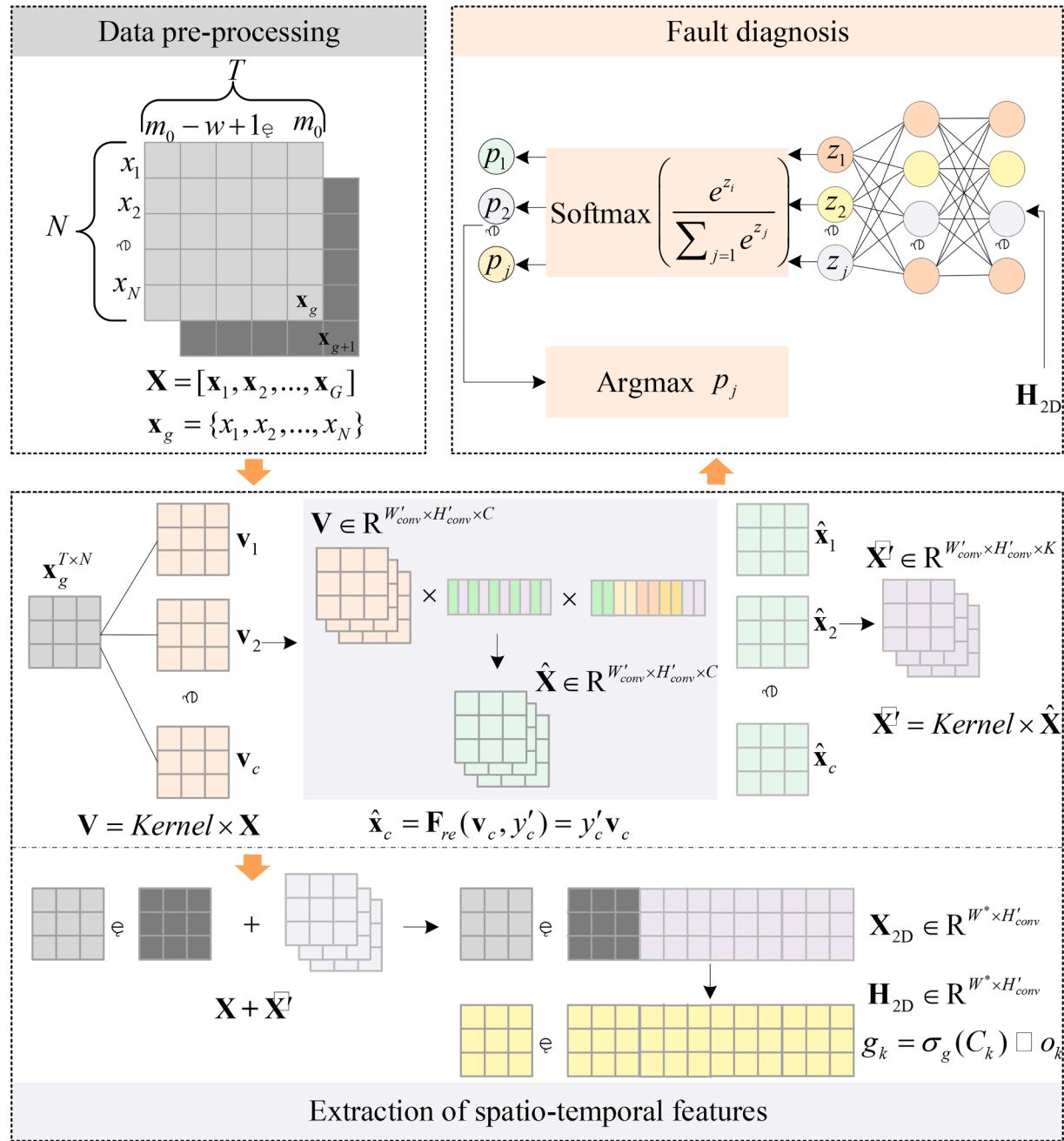


Fig. 8. Flow chart of the proposed 3DSE-CNN-2DLSTM models.

$$f_k = \sigma_s(R_f[g_{k-1}, x_k] + b_f) \quad (7)$$

$$i_k = \sigma_s(R_i[g_{k-1}, x_k] + b_i) \quad (8)$$

$$o_k = \sigma_s(R_o[g_{k-1}, x_k] + b_o) \quad (9)$$

$$\tilde{C}_k = \sigma_g(R_C[g_{k-1}, x_k] + b_C) \quad (10)$$

$$C_k = f_k \odot C_{k-1} + i_k \odot \tilde{C}_k \quad (11)$$

$$g_k = \sigma_g(C_k) \odot o_k \quad (12)$$

where σ_s is the sigmoid activation function; σ_g is the hyperbolic tangent activation function; the forgetting gate f_k is obtained from the weight matrix R_f of the input initial state $[g_{k-1}, x_k]$ at moment k and the bias parameter b_f through σ_s ; the input gate i_k is obtained by updating the

- (7) weight matrix R_i and the bias parameter b_i of the input initial state $[g_{k-1}, x_k]$ at moment k through σ_s ; the output gate o_k is obtained from the weight matrix R_o of the input initial state $[g_{k-1}, x_k]$ at the time k and the bias parameter b_o ; the new vector \tilde{C}_k is obtained from the weight matrix R_C of the input initial state $[g_{k-1}, x_k]$ at the time k and the bias parameter b_C through σ_g ; the updated cell state C_k is a combination of two types of information, an inputting gate i_k and a forgetting gate f_k ; an updated hidden state g_k is obtained i_k and o_k through σ_g ; \odot denotes Hadamard product, a vector of elements multiplied by each other.

3.3. Proposed 3DSE-CNN-2DLSTM model

CNN extracts local features through a layer-by-layer structure. However, when training high-dimensional data, CNN usually suffers from an over-fitting problem. To solve the overfitting problem, several

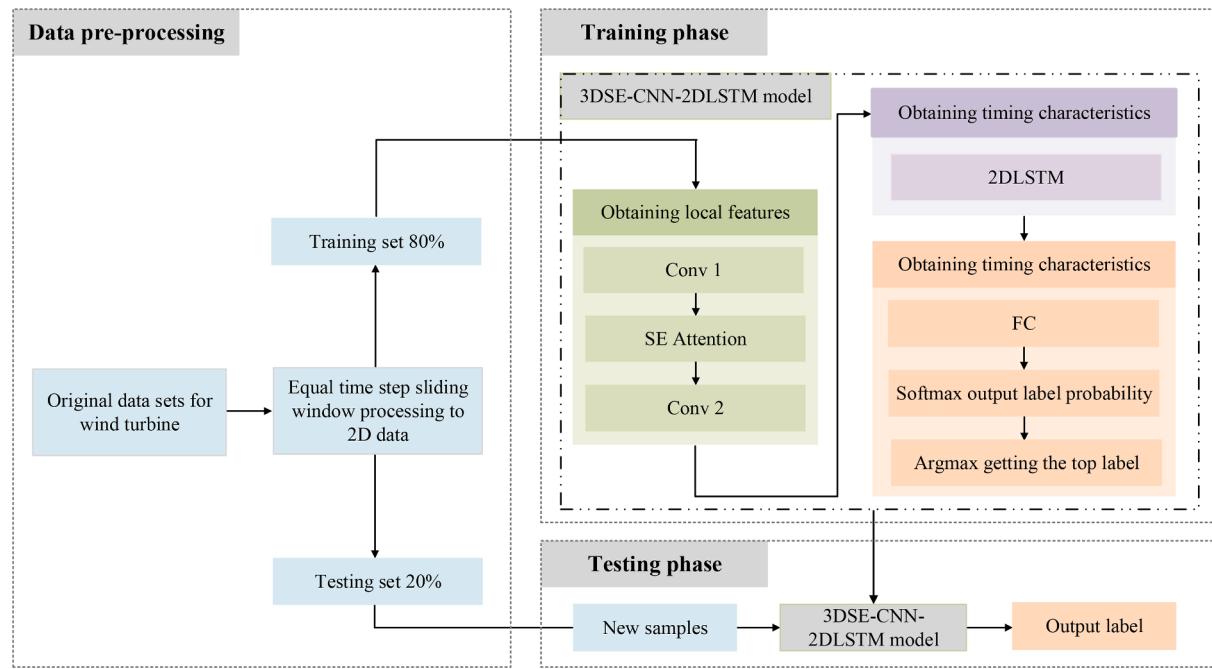


Fig. 9. WT fault detection steps for the proposed 3DSE-CNN-2DLSTM model.

Table 5

Sample size statistics by category.

Data set	Fault62	Fault60	Fault228	Fault80	Fault9	Fault60-228	Fault62-80	No-fault	Total samples
Training set	129	6	40	61	35	9	80	25,639	25,999
Testing set	33	2	10	16	9	3	21	6410	6504

Table 6

Hybrid 3DSE-CNN-2DLSTM model parameter setting.

	Layers	Designation of layer	Output	Kernel size/ Step/ Zero-filling
3DSE-CNN-2DLSTM	Layer 1	Conv2d_1 (Conv2D)	(10,22,22)	(3,3)/1/1
	Layer 2	SE attention	(10,22,22)	–
	Layer 3	Conv2d_2 (Conv2D)	(20,22,22)	(3,3)/1/1
	Layer 4	Trans	(22,440)	–
	Layer 5	LSTM_1 (LSTM)	(22,462)	–
		LSTM_2 (LSTM)	(22,462)	–
		LSTM_3 (LSTM)	(22,462)	–
	Layer 6	Dense_1 (Dense)	(256)	–
Layer 7	Dropout	(256)	–	
	Dense_2 (Dense)	(8)	–	

Table 7

Comparison of experimental model parameter settings.

Model	Parameter
CNN	Total count of iterations = 200, number of channels = [10,20,35], dropout = 0.2
LSTM	Total count of iterations = 200, dropout = 0.1
GRU	Total count of iterations = 200, dropout = 0.1
TCN	Total count of iterations = 200, number of channels = [64, 16, 4, 1], dropout = 0.1
3DSE-CNN-2DLSTM	Total count of iterations = 200, number of channels [10,20], dropout = 0.2

Table 8

Performance evaluation formula.

Parameter matrix	Formula
Accuracy	$Acc = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}$
Each class of precision	$Precision_i = \frac{TP_i}{TP_i + FP_i}$
Each class of recall	$Recall_i = \frac{TP_i}{TP_i + FN_i}$
Each class of f1-score	$F1 - score_i = 2 \times \frac{Precision_i \times Recall_i}{Precision_i + Recall_i}$
Multiclassification macro-precision	$Macro - pre = \frac{\sum_{i=1}^k Precision_i}{k}$
Multiclassification macro-recall	$Macro - re = \frac{\sum_{i=1}^k Recall_i}{k}$
Multiclassification macro-F1	$Macro - F1 = \frac{\sum_{i=1}^k F1 - score_i}{k}$

research methods have been proposed, including dropout (Mohine et al., 2022), batch normalization (Segu et al., 2023), and data enhancement (Niu et al., 2022). In contrast, LSTM performs well when training high dimensional data and time series owing to that unique structure. However, in terms of feature extraction, LSTM may lose some distinct features compared to CNN. Consequently, the advantages and disadvantages of the two models are utilized to select the appropriate model to improve performance when choosing a model. Meanwhile, when employing CNN, several methods are applied to mitigate the overfitting phenomenon.

When building a DL model, choosing model depth is a very critical hyperparameter selection problem. While increasing the depth of the model has the potential to improve model learning, this is not always the case. The deep models require an additional amount of computational

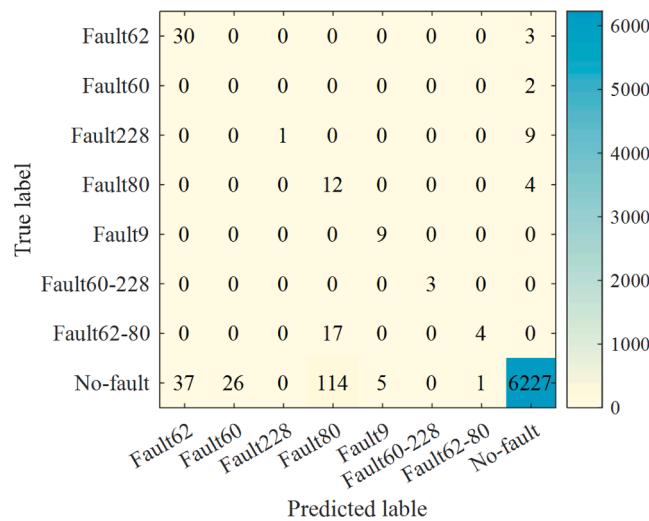


Fig. 10. Confusion matrix of 3DSE-CNN-2DLSTM.

resources to train and deploy. The deep models are prone to problems with gradient disappearance or gradient explosion, which can affect the effectiveness of the model.

Therefore, to save computational resources and obtain excellent model results, the best option is to utilize as few layers as possible for a task under study. Excessively numerous layers may result in problems regarding over-fitting. Accordingly, this study considers task

requirements, computational resources, and model effectiveness when selecting the model depth to find the optimal balance.

To extract high-dimensional features and avoid overfitting, the proposed 3DSE-CNN-2DLSTM model is trained and tested in two stages. Two layers of CNN, three layers of LSTM, and two layers of full connectivity are designed in the proposed 3DSE-CNN-LSTM. Fig. 3 indicates the model framework. With the designed 3DSE-CNN-2DLSTM model, the proposed 3DSE-CNN-2DLSTM model can obtain high dimensional features with appropriate accuracy and avoid overfitting during the training phase. In addition, dropout is deployed to alleviate the overfitting problem. The specific flow of the proposed 3DSE-CNN-2DLSTM model (Fig. 8) includes three key components. In part 1, the data pre-processing is performed with a sliding window approach. In part 2, spatiotemporal fusion features are extracted by the proposed 3DSE-CNN and 2DLSTM models. Finally, fault detection is carried out with softmax and argmax through the fully connected layer.

Table 9
Performance indicators for all models.

Model	Accuracy	Macro-precision	Macro-recall	Macro-F1
CNN	0.967	0.642	0.560	0.492
LSTM	0.898	0.551	0.571	0.472
GRU	0.946	0.472	0.497	0.420
TCN	0.857	0.514	0.666	0.484
3DSE-CNN-2DLSTM	0.967	0.621	0.615	0.501

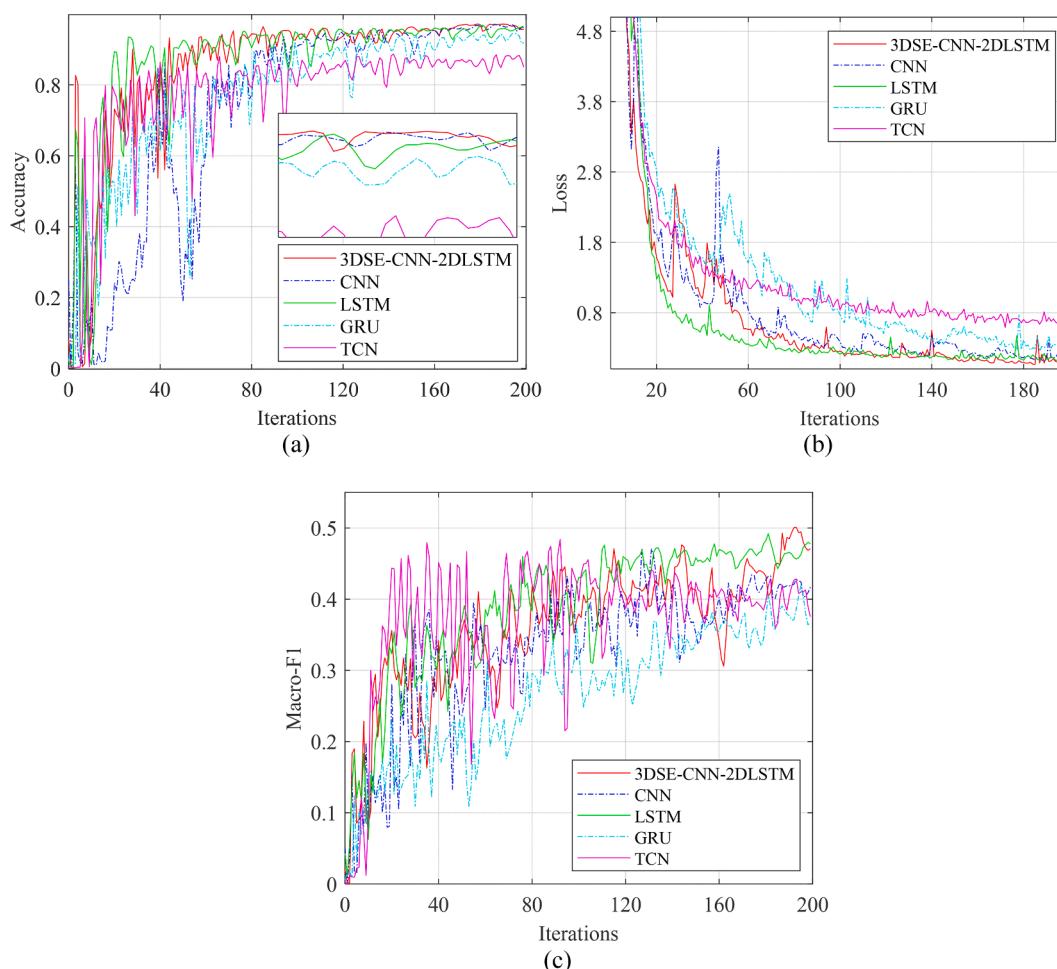


Fig. 11. Recognition results for 200 iterations of each model (a) accuracy curve; (b) loss curve; (c) macro-F1 curve.

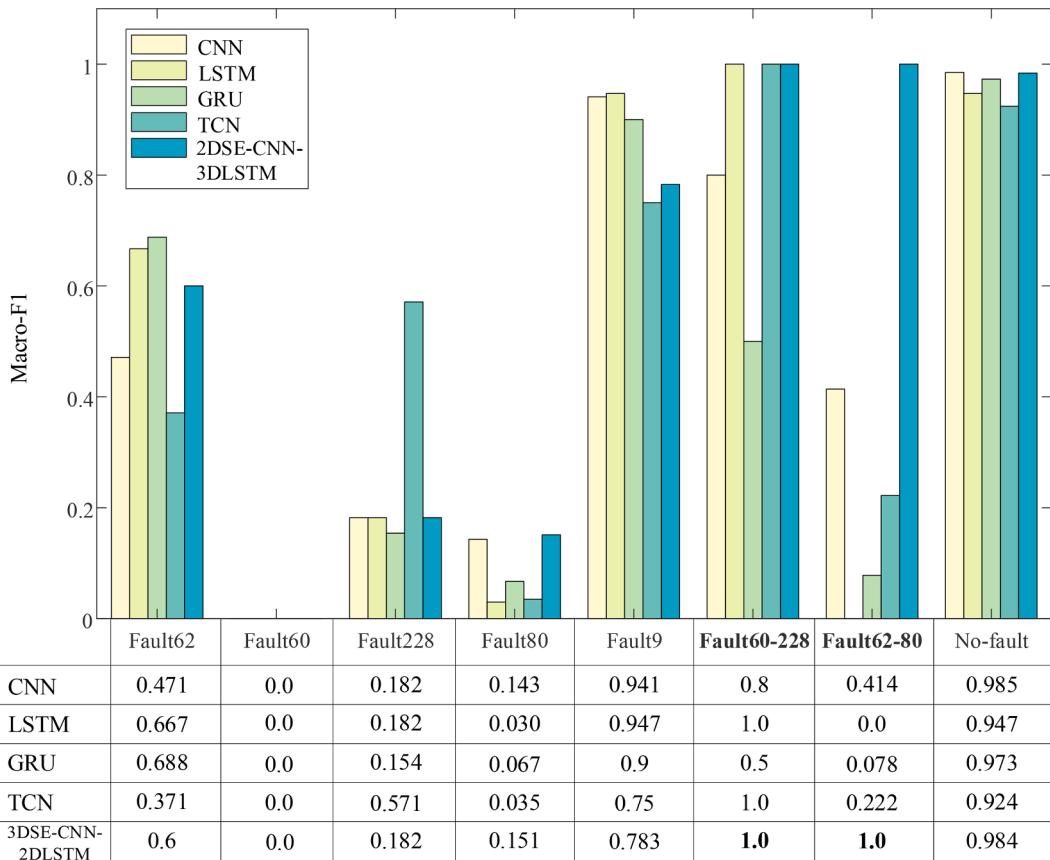


Fig. 12. Macro-F1 of CNN, LSTM, GRU, TCN, and 3DSE-CNN-2DLSTM on each fault class.

3.4 wt. fault detection based on the proposed 3DSE-CNN-2DLSTM models

The 3DSE-CNN and 2DLSTM of 3DSE-CNN-2DLSTM combine the spatial and temporal information in WT SCADA data to mine comprehensive and accurate spatiotemporal fusion information. Specifically, firstly, the proposed 3DSE-CNN model extracts local features from the WT SCADA data, namely capturing different spatial features at various points in time. The SE attention selects key features to increase the weighting of key features in the representation. Then, the proposed 2DLSTM model extracts the spatiotemporal fusion features. Finally, the class label with the highest probability is obtained by argmax. The WT fault detection steps and pseudo-code for the proposed 3DSE-CNN-2DLSTM model are given in Fig. 9 and algorithm 1, respectively.

Algorithm 1: Pseudocode for the proposed 3DSE-CNN-2DLSTM model.

- 1: Initialization parameters (the proposed 3DSE-CNN-2DLSTM model: learning rate, number of channels, layers and neurons, step size and window length of sliding window, number of iterations, batch)
- 2: At system runtime, execute
- 3: Input SCADA data and process with the method in Fig. 5 to generate a 2D original image \mathbf{X}
- 4: 3D image feature $\tilde{\mathbf{X}}$ of SCADA system is obtained from the proposed 3DSE-CNN model in Fig. 3 with equation (1)-(6)
- 5: From the proposed transformation layer in Fig. 3, the 3D image feature data $\tilde{\mathbf{X}}$ and the 2D original image \mathbf{X} are transformed into 2D image data with a single-time step.
- 6: Proposed 2DLSTM model from Fig. 3 is extracted from the temporal features with equations (7)-(11) and fused with the spatial features to form the final fused features
- 7: Softmax obtains fault class probabilities
- 8: Fault category label is output by argmax

The proposed 3DSE-CNN-2DLSTM model is utilized to detect the compound fault of WT. Compared to existing DL-based data-driven methods, the data preprocessing method achieves rapid dimensionality reduction

for data and retains spatial and temporal information on the occurrence of faults. The complexity of WT systems can have the following drawbacks: (1) the SCADA data is highly dimensional and contains highly correlated features; therefore, the detection of compound fault is not easily and accurately carried out; (2) faults occur usually in a cumulative manner over time; (3) most detected methods for compound faults are based on specific problems and rely on a large amount of a priori knowledge.

The proposed 3DSE-CNN-2DLSTM model can effectively solve the drawbacks, summarized below: (1) original data is integrated into the 2D original image data by the sliding window method; the original 2D image data retains spatial and temporal information, providing sufficient fault detection information for the proposed 3DSE-CNN-2DLSTM model; (2) a novel 3DSE-CNN-2DLSTM model can extract spatiotemporal fusion features to solve the composite fault detection problem of WTs; (3) the proposed 3DSE-CNN-2DLSTM model is based on a data-driven model. Hence, the proposed 3DSE-CNN-2DLSTM model can work with other compound fault problems.

4. Experimental verification

To evaluate performance with 3DSE-CNN-2DLSTM, a DL framework is built with Python 3.9.7 and Pytorch on a model V100-SXM2-32 GB GPU graphics card. In this study, the 3DSE-CNN-2DLSTM model is trained and tested based on the WT SCADA data description and data labels in Section 2. The parameters of the proposed 3DSE-CNN-2DLSTM model are set through a training process. According to the evaluation metrics, the 3DSE-CNN-2DLSTM model is compared with other classical models for experiments. The 3DSE-CNN-2DLSTM model SE module and overlaid original images are discussed.

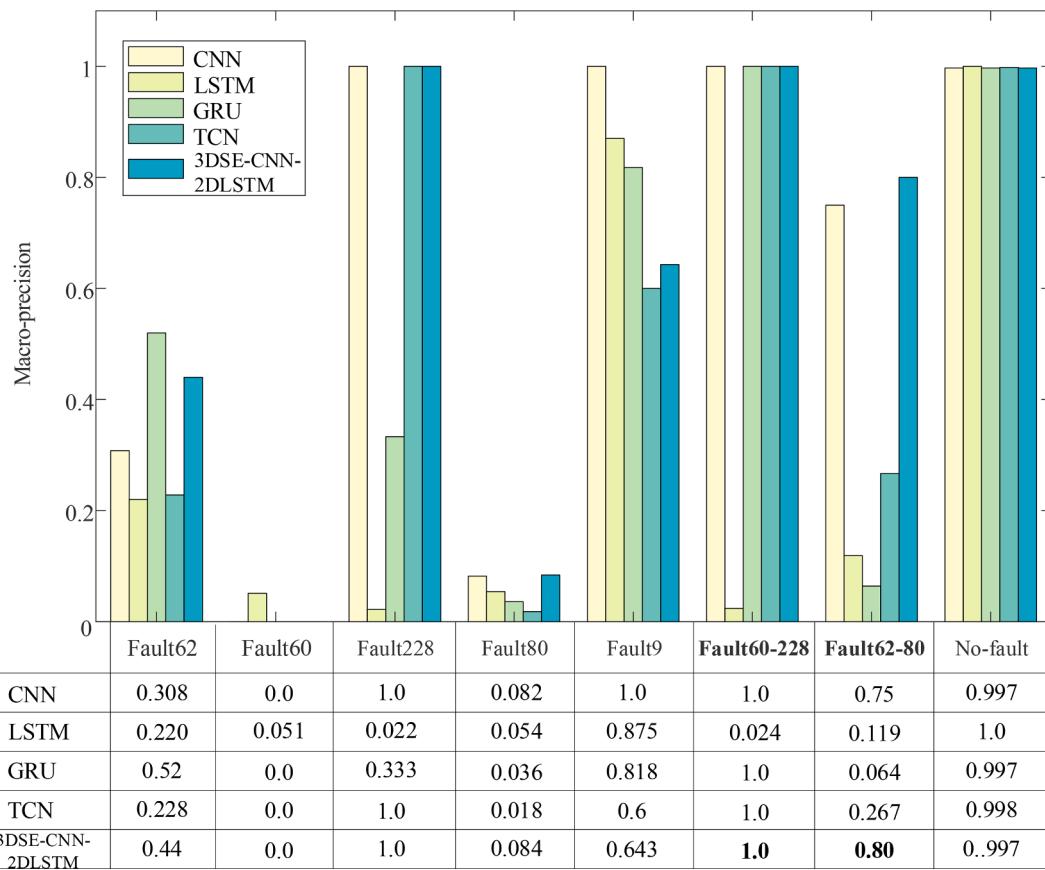


Fig. 13. Macro-precision of CNN, LSTM, GRU, TCN, and CNN-LSTM on each fault class.

4.1. Experimental setup

The 22 variables of the selected WT SCADA data reflect the WT operating status. Operating categories for WTs include seven fault categories and one normal state. A total of 80 % with SCADA data serves for training. Meanwhile, random noise is added to the training set for data enhancement. A further 20 % with SCADA data served for testing. Parameters for the sliding window method are set to: $T = 22$, $\beta = 1$. Hence, the training set of 2D image samples obtained after sliding window processing is 25,999 and the test set is 6504. Table 5 shows the training and test sets of the 2D image samples for each category. The single fault categories fault62, fault60, fault228, fault80, and fault9 denote feeding fault, excitation fault, air-cooling fault, mains fault, and generator fault, respectively. The compound fault categories fault60-228, and fault62-80 denote feeding and excitation fault, mains and air-cooling fault, respectively.

The 3DSE-CNN-2DLSTM parameters are significant for performance. Particularly, a learning rate, the proposed 3DSE-CNN-2DLSTM model tests the model according to exponentially decreasing values of different learning rates. Macro-F1 is highest when the number of iterations is 200 and the learning rate reaches 0.001. Therefore, both the Adam optimizer and the FL function are set to a 0.001 learning rate. After the SCADA data preprocessing, the 3DSE-CNN-2DLSTM model starts the convolution operation. In the 3DSE-CNN, the first convolution layer has an input channel of one and an output channel of ten. In the SE block, the input and compression multiplier are ten and two, respectively. The second convolution layer has an input channel of ten and an output channel of twenty. The proposed 2D LSTM model is a three-layer LSTM with the same parameter settings. The sizes for the input and output 2D images are (22,440) and (22,462), respectively. Finally, the fully connected layer expands the 2D image features and outputs as 256. The classification outcomes are output by both softmax and argmax functions. The

parameters of the 3DSE-CNN-2DLSTM model are the optimal ones obtained after several training sessions. Table 6 shows the other parameter settings of the proposed 3DSE-CNN-2DLSTM model.

In this study, several classical DL models are employed for comparative experiments of WT compound fault detection. CNN models belong to the fault detection methods of feature learning (Huang et al., 2023). TCN hierarchical capture of spatiotemporal and temporal information (Luo et al., 2022), solves fault problems (Cheng et al., 2023). GRU (Wang et al., 2022) and LSTM (Zhang et al., 2022) are outstanding in extracting fault temporal information (Ma et al., 2022). Table 7 shows the relevant parameters for the comparison models.

(1) All models have the same learning rate setting (e.g., 3DSE-CNN-2DLSTM, CNN, LSTM, GRU, and TCN). The depth of the models is set approximately equal to the complexity of each model.

(2) To achieve single-label multi-categorization, a fully connected layer with the same parameters is added at the end of all models.

This study introduces an FL function. The FL function assigns different weights according to the sample size of different classes, which helps to balance the training process and prevents the model from being biased towards the majority class. 3DSE-CNN-2DLSTM performance is improved in the compound fault detection of WT (Xiang et al., 2022). The FL equation is given below:

$$FL(p_i) = -\alpha_i(1-p_i)^\gamma \log(p_i) \quad (13)$$

where α_i and γ are hyperparameters, given as 0.25 and 2, respectively; p_i is a likelihood that the model predicts a positive sample in category i th.

4.2. Evaluation indicators

For a full evaluation of the compound fault detection performance with 3DSE-CNN-2DLSTM, this subsection introduces several types of

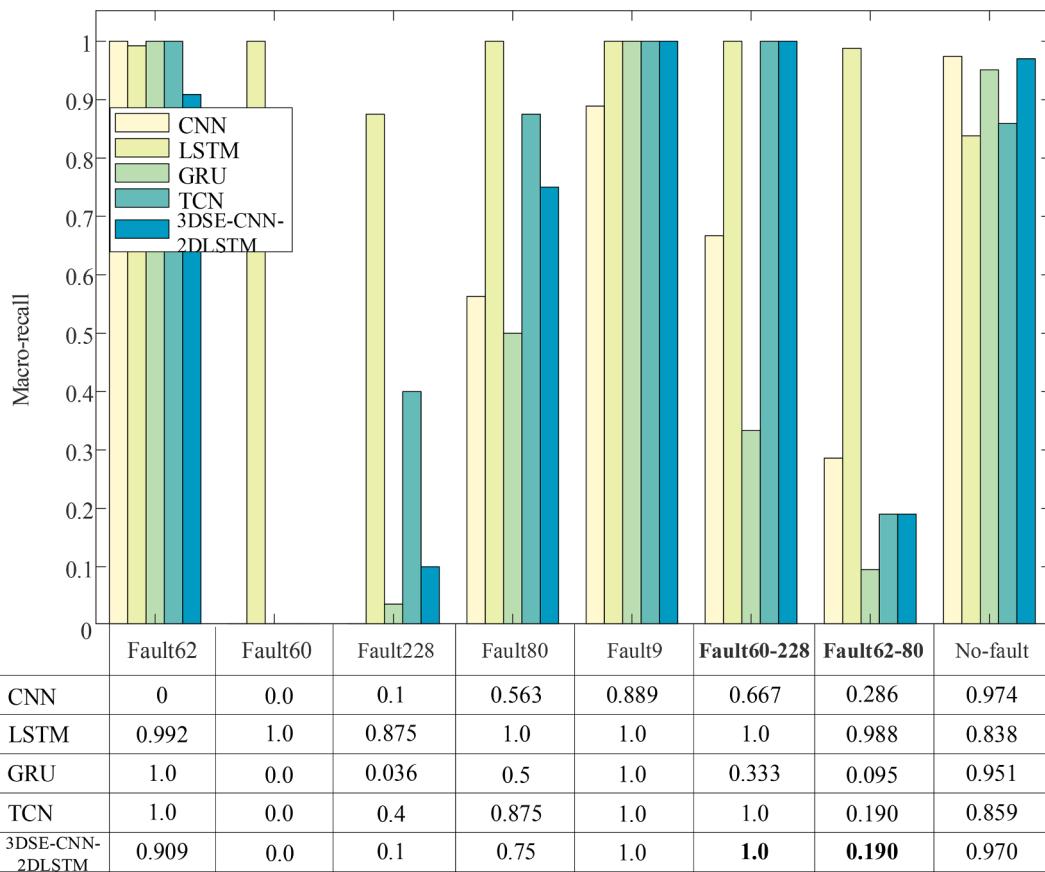


Fig. 14. Macro-recall of CNN, LSTM, GRU, TCN, and CNN-LSTM on each fault class.

metrics. Macro-F1 is a metric for evaluating the performance of multi-category classifiers that measures the balance of predictive precision and recall across categories and provides a fair assessment of data with unbalanced categories. A high macro-F1 indicates that the model achieves superior equilibrium performance across multiple categories. Table 8 gives indicator formulas. TP_i represents correctly labeled positive samples. FP_i represents mislabeled positive samples. TN_i represents correctly labeled negative samples. FN_i represents mislabeled negative samples. Furthermore, k represents the total categories, and i represents a category.

4.3. Detection results and comparison

The effectiveness generated from the 3DSE-CNN-2DLSTM for detecting compound faults is verified for its effectiveness with the confusion matrix (Fig. 10). The confusion matrix analysis revealed that (1) the fault detection result for the fault60-228 compound category is perfect; (2) the fault62-80 compound category has 17 samples incorrectly detected as fault80 and 1 sample incorrectly detected as no-fault. The main reason for this result in this study analysis is the low feature distinction between the single category fault80 and the compound category fault62-80.

At the 200th iteration, the best set of data from each of the model training results is selected. The detection results of the different models are presented in Fig. 11, namely the CNN, LSTM, GRU, TCN, and 3DSE-CNN-2DLSTM models. The results of the study show that: (1) as shown in Fig. 11(a), the accuracy of all models increases as the number of iterations increases and plateaus after 150 iterations. At 200 iterations, the proposed 3DSE-CNN-2DLSTM, CNN, and LSTM show higher accuracy than the other models, demonstrating the advantages of 3DSE-CNN and CNN in extracting spatial features and 2DLSTM and LSTM in extracting temporal features. (2) Fig. 11(b) shows that although LSTM

converges faster than the proposed 3DSE-CNN-2DLSTM in the early stage, the proposed 3DSE-CNN-2DLSTM converges faster at the later stage. (3) According to Fig. 11(c), as the number of iterations increases, the macro-F1 of TCN fluctuates more than other models; the overall macro-F1 of GRU is the lowest compared to other models; the macro-F1 values of LSTM and the proposed 3DSE-CNN-2DLSTM are higher compared to other models.

While both CNN and LSTM show better performance during training, the proposed 3DSE-CNN-2DLSTM had the best combined performance during testing. The overall ability of all models to correctly classify (Table 9) shows that the 3DSE-CNN-2DLSTM model achieved the highest values in terms of accuracy and macro-F1 metrics, with both macro-precision and macro-recall reaching above 0.6. Consequently, the proposed DSE-CNN-2DLSTM performs best at the time of testing.

To evaluate the effectiveness of 3DSE-CNN-2DLSTM in capturing temporal characteristics from SCADA data, a comparative analysis is conducted between the 3DSE-CNN-2DLSTM and two other models, namely GRU and LSTM. The experiment aims to assess 3DSE-CNN-2DLSTM compared to other commonly employed models in capturing and analyzing the temporal patterns of SCADA data. The performance metrics for macro-F1, macro-precision, and macro-recall for the eight categories corresponding to each model (Figs. 12-14) show that (1) the proposed 3DSE-CNN-2DLSTM, GRU, and LSTM all have excellent classification performance on single faults classes; (2) when detecting the compound faults fault60-228 and fault62-80, the macro-F1 of the LSTM are 1 and 0; the macro-F1 of the GRU are 0.5 and 0.078; the macro-F1 of the 3DSE-CNN-2DLSTM are 1 and 1; (3) when detecting the compound faults fault60-228 and fault60-80, compared with GRU and LSTM, the proposed 3DSE-CNN-2DLSTM model has relatively higher macro-precision and macro-recall. The confusion matrix (Fig. 10, Fig. 15(a), and Fig. 15(c)) further demonstrates the number of correct and predicted labels corresponding to each category in each model.

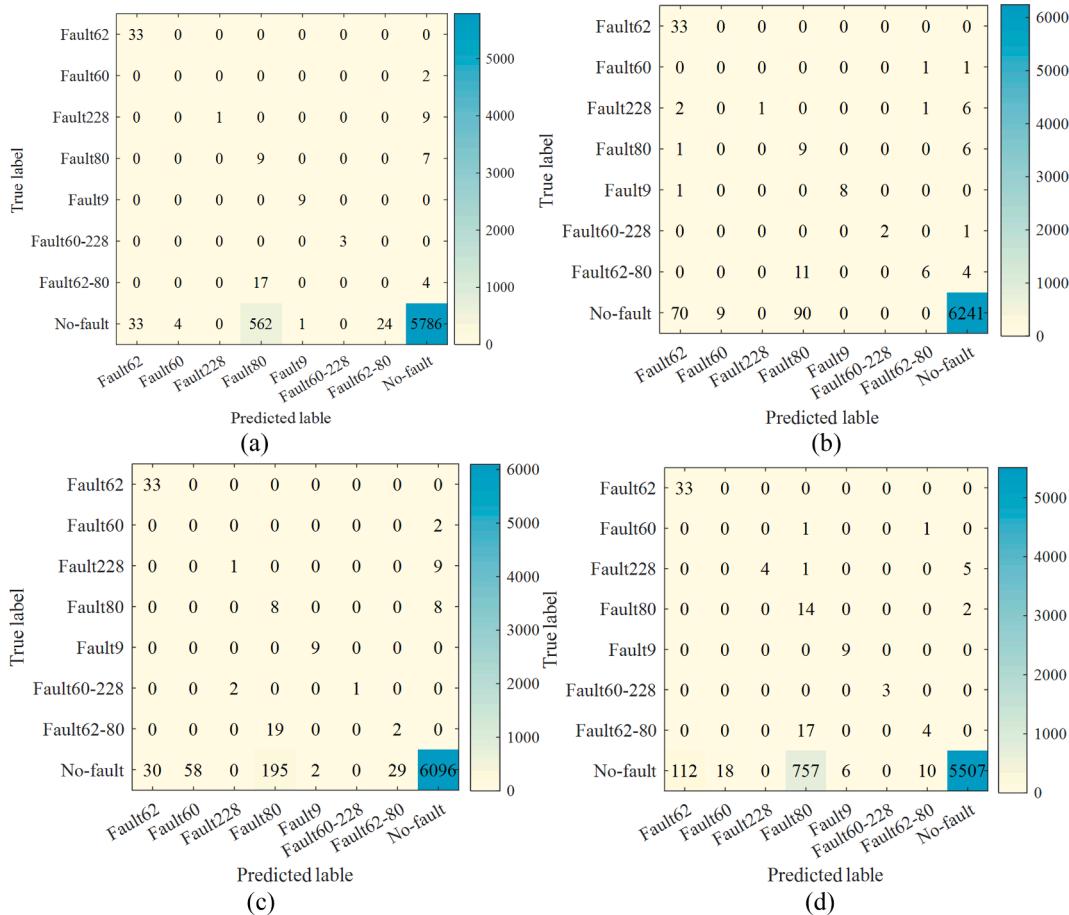


Fig. 15. Confusion matrix with different methods: (a) LSTM; (b) CNN; (c) GRU; (d) TCN.

Table 10

Accuracy comparison of existing and proposed 3DSE-CNN-2DLSTM models.

Model	Fault categories	Accuracy
ADLC (Zhu and Song, 2022)	Normal and faulty	0.996
GPC (Li et al., 2019)	Fault-free, feeding fault, air-cooling, excitation, generator, mains	0.938
STMNN (Zhu and Zhang, 2021)	Normal, feeding fault, excitation, air-cooling, mains, generator	0.973
SVM (Leahy et al., 2016)	Fault / no-fault	0.994
3DSE-CNN-2DLSTM	No-fault, feeding fault, air-cooling, excitation, generator, mains, feeding and excitation fault, mains and air-cooling fault	0.967

Table 11

Results of SE modules discussed in eight group models.

Model	Accuracy	Macro-precision	Macro-recall	Macro-F1
SECNN- SECNN-LSTM	0.969	0.535	0.462	0.458
CNN-SECNN-LSTM	0.976	0.556	0.419	0.440
CNN-CNN-LSTM	0.921	0.499	0.630	0.432
SECNN-CNN-LSTM	0.967	0.621	0.615	0.501
SECNN-SECNN-SECNN-LSTM	0.966	0.651	0.469	0.483
SECNN-SECNN-CNN-LSTM	0.947	0.505	0.453	0.409
SECNN-CNN-CNN-LSTM	0.923	0.566	0.643	0.494
CNN-CNN-CNN-LSTM	0.863	0.498	0.625	0.447

The effectiveness of 3DSE-CNN-2DLSTM in capturing spatial relationships between variables in SCADA data is evaluated by comparing with both CNN and TCN models. The eight categories of macro-F1, macro-precision, and macro-recall corresponding to each model (Figs. 12–14) show that (1) the proposed 3DSE-CNN-2DLSTM, CNN, and TCN all have excellent classification performance on single faults classes; (2) when detecting the compound faults fault60-228 and fault62-80, the macro-F1 of CNN is 0.8 and 0.414; the macro-F1 of TCN is 1 and 0.222; the macro-F1 of 3DSE-CNN-2DLSTM is 1 and 1; (3) the proposed 3DSE-CNN-2DLSTM model has excellent performance in detecting compound fault fault60-228 compared to GRU and LSTM. Meanwhile, the proposed 3DSE-CNN-2DLSTM has a macro-F1 value of 1 when detecting compound fault fault60-80. By employing confusion matrices (Fig. 10 and Fig. 15(b), (d)), the classification of 3DSE-CNN-2DLSTM, CNN, and TCN models are analyzed in detail.

In summary, the performance of the proposed 3DSE-CNN-2DLSTM model in this study is not significantly improved when performing single fault detection though. However, when detecting the compound fault types (fault60-228 and fault62-80), the proposed 3DSE-CNN-2DLSTM achieves the highest values of macro-F1, both of which are 1 (columns 7 and 8 in the sixth row of Fig. 12). Macro-precision reaches the highest values of 1 and 0.8 (row 6, columns 7 and 8 of Fig. 13). The values for macro-recall are 1 and 0.19 (row 6, columns 7 and 8 of Fig. 14). Hence, the proposed 3DSE-CNN-2DLSTM extracts deep spatiotemporal fusion features; the compound fault detection performance is excellent. Particularly noteworthy is that the macro-F1 of fault60 is 0. Explaining the phenomenon (Table 5), fault60 has merely 2 training samples with few training samples.

The experimental outcomes indicate that (1) the 3DSE-CNN-2DLSTM converges quickly during training and validates the efficiency of

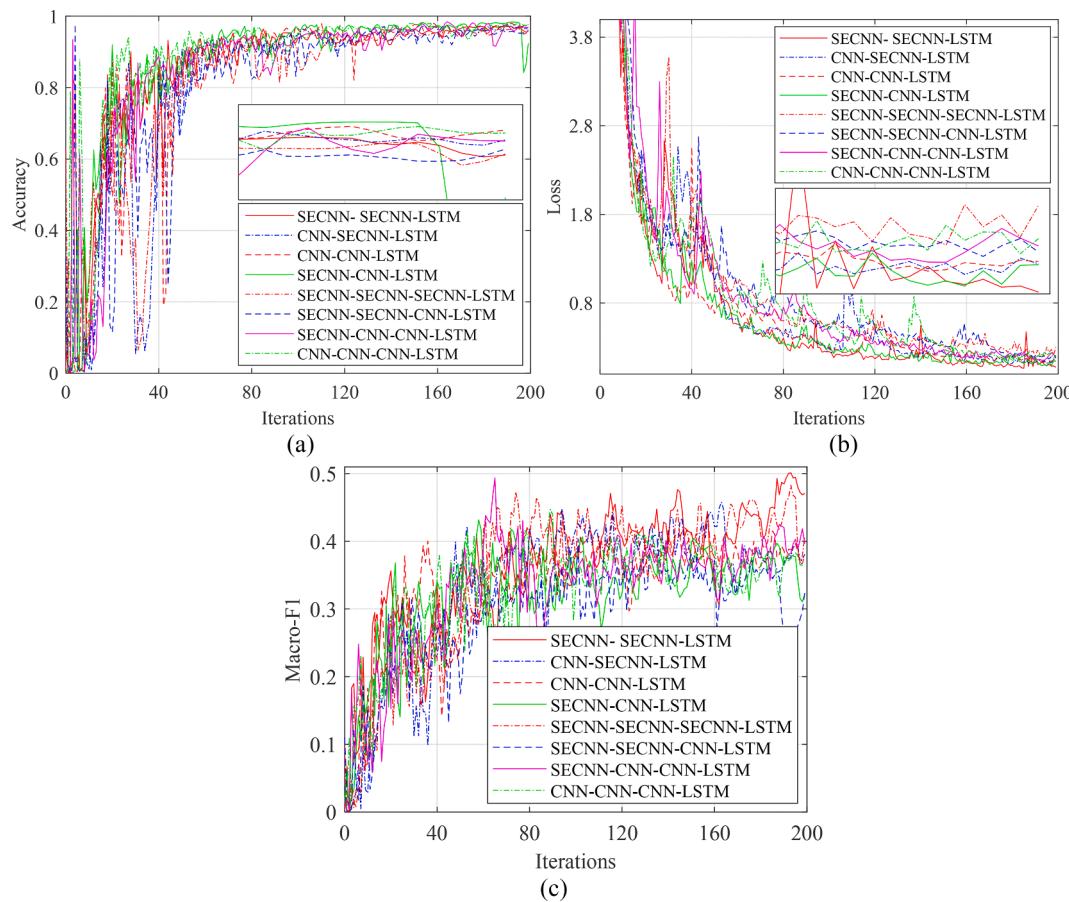


Fig. 16. Recognition results of SE modules for 200 iterations in eight group models: (a) accuracy; (b) loss curve; (c) macro-F1.

compound fault detection; (2) compared with GRU and LSTM, the proposed 3DSE-CNN-2DLSTM model can mine deep temporal features and spatiotemporal fusion features; the proposed 3DSE-CNN-2DLSTM model performs excellently in extracting spatial features compared with CNN and TCN; (3) the proposed 3DSE-CNN-2DLSTM model improves the classification performance of WT compound fault detection.

To evaluate the composite fault detection capability of 3DSE-CNN-2D LSTM, this study compares the overall accuracy of existing models (i.e., ADLC, GPC, STMNN, and SVM) applied to the same SCADA dataset with 3DSE-CNN-2D LSTM (Table 8). The analysis reveals that the fusion of 3DSE-CNN and 2DLSTM models has a significant impact on the extraction of spatiotemporal fusion features. In terms of accuracy (Table 10) shows the comparison between the existing model and the proposed 3DSE-CNN-2DLSTM model. The fault accuracy of the automatic data label calibration (ADLC) based binary classification model was 0.996 (Zhu and Song, 2022). The accuracy of a single multi-classification fault detection based on a Gaussian process classifiers (GPC) model was 0.938 (Li et al., 2019). The accuracy of the single multi-classification fault detection based on the spatiotemporal multi-scale neural network (STMNN) model was 0.973 (Zhu and Zhang, 2021). The fault accuracy of the SVM-based binary classification model was 0.994 (Leahy et al., 2016). The accuracy of each model is high for either binary classification or single multi-classification with the same SCADA dataset. However, none of these models take into account the presence of specific compound faults in the SCADA data. The proposed 3DSE-CNN-2DLSTM model considers some specific compound fault categories with an accuracy of 0.967. Accordingly, the proposed 3DSE-CNN-2DLSTM model enables some specific compound fault detection.

4.4. Discussions

To verify the validity of the SE module and overlay the original image in the proposed 3DSE-CNN-2DLSTM model, a comparison is discussed in this section.

4.4.1. SE module

In this study, two-layer CNN and three-layer CNN are designed in the proposed 3DSE-CNN-2DLSTM model. A SE module is added to different CNN layers. In this subsection, the SE module is added to the CNN layer to form an eight-group model. The locations of the added SE modules are given in Table 11. The eight group models designed in this subsection have the same structure and parameter settings except for the SE module and the number of CNN layers.

The 3DSE-CNN-2DLSTM model proposed in this study is a SECNN-CNN-LSTM model, which adds an SE module to the first layer of the two-layer CNN structure. Fig. 16 shows the training and testing results for eight groups of models over 200 iterations. Comparisons between the accuracy of the eight groups of models (Fig. 16(a)) show that the accuracy of all models leveled off after 50 iterations. The loss function curves for the eight groups of models (Fig. 16(b)) show that all eight groups of models converge as the number of iterations increases; the SECNN-CNN-LSTM model converges faster than the other seven groups of models and has the smallest final loss value. The macro-F1 of the eight groups of models (Fig. 16(c)) shows that the macro-F1 of the models with the SE module added is higher than the models without the SE module added. In particular, the proposed SECNN-CNN-LSTM model performed best after 200 iterations in this study. In summary, the SE module allows for the screening of important feature information and enhances spatial feature extraction. The contribution of different channel information in convolutional layers is utilized differently for fault

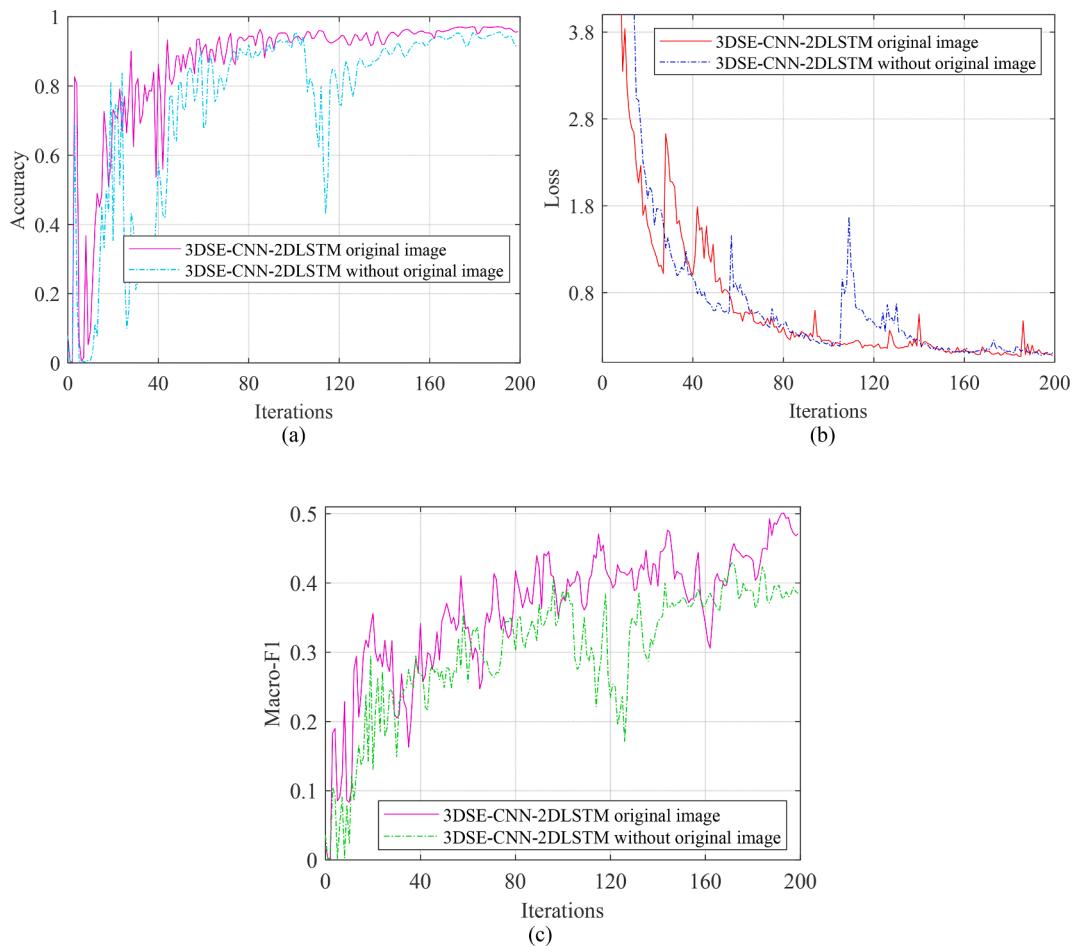


Fig. 17. Results after 200 iterations of the original image: (a) accuracy; (b) loss; (c) macro-F1.

Table 12

Discussion of overlaying the original image.

Method	Accuracy	Macro-precision	Macro-recall	Macro-F1
3DSE-CNN-2DLSTM without original image	0.948	0.493	0.522	0.429
3DSE-CNN-2DLSTM original image	0.967	0.621	0.615	0.501

detection. Consequently, the SE module is added to the CNN layer in the proposed 3DSE-CNN-2DLSTM model in this study.

4.4.2. Overlay the original image

A discussion study is conducted to validate the effectiveness of overlaying original images into a 2D LSTM model. The experimental group 3DSE-CNN-2DLSTM overlaid with original images. The control group 3DSE-CNN-2DLSTM without original images is overlaid.

The detected results discussed with overlaid the original image (Fig. 17) show that (1) the accuracy curve, loss curve, and macro-F1 curve corresponding to the control group are encompassed broadly by the experimental group, respectively; (2) the 3DSE-CNN-2DLSTM original image model shows more stable accuracy and faster convergence concerning the control group; (3) the 3DSE-CNN-2DLSTM original image model outperformed the control group in terms of performance metric macro-F1. Furthermore, a comparison of evaluation indicators between the two models for the testing set (Table 12) shows that the experimental group model improved in all performance metrics, with an increase of 0.128, 0.093, and 0.072 for macro-precision, macro-recall,

and macro-F1, respectively. The experimental and control groups demonstrate the advantages of overlaying the original image in both training and testing.

The above two groups of discussions demonstrate that (1) among eight groups of models, the SECNN-CNN-LSTM can select channel information, extracts spatial features, and has excellent classification performance when performing fault detection; (2) overlaying the original image at the input of the 2DLSTM model preserves the original information and enhances feature representation when extracting temporal features from the 2DLSTM; (3) the overall performance of 3DSE-CNN-2DLSTM is improved by adding SE module to the first CNN layer of the two-layer CNN and overlaying the original image.

In the experiments and discussions, the proposed 3DSE-CNN-2DLSTM model is trained and tested. To verify that the proposed 3DSE-CNN-2DLSTM model mine spatiotemporal feature information and extract spatiotemporal fusion features, the WT compound fault detection problem is solved. The drawbacks of this study are summarized as follows, (1) although the FL function is employed to train the model and alleviate the data imbalance problem, some single fault classes do not have high classification performance in the model in multi-classification detection; (2) a longer time is required to label the categories of the training samples before training the model.

5. Conclusions

In this study, a WT composite fault detection model with a fusion between 3DSE-CNN and 2DLSTM is proposed by considering spatial information and time series in SCADA data. Experiments are conducted on a 3GW SCADA dataset. The 3DSE-CNN-2DLSTM has high accuracy

for composite fault detection. The main contributions of 3DSE-CNN-2DLSTM are summarized as follows.

- (1) This study utilizes sliding window techniques and transformation techniques for data preprocessing to convert SCADA data to 1D vector data to 2D image data, which prepares the proposed 3DSE-CNN-2DLSTM model for mining deep spatiotemporal information.
- (2) The proposed 3DSE-CNN-2DLSTM model has high detected performance. The proposed 3DSE-CNN-2DLSTM model is designed in the proposed 3DSE-CNN model, which screens the heterogeneous information within each channel. The experiments show that the proposed 3DSE-CNN-2DLSTM model has the highest macro-F1 values among the eight groups of models with SE modules added in different layers. Meanwhile, the 2DLSTM model is overlaid with the original image at the input. The macro F1 value of the experimental group regarding the superimposed original image setting is 0.072 higher than that of the control group. Moreover, the proposed 3DSE-CNN-2DLSTM model outperforms CNN, LSTM, GRU, and TCN models in detecting specific compound fault categories. While for Fault60-228, the LSTM, TCN, and 3DSE-CNN-2DLSTM models have the same or very similar fault detection performance, for Fault62-80, the proposed macro-F1 of the 3DSE-CNN-2DLSTM model exhibits excellent performance.
- (3) The proposed 3DSE-CNN-2DLSTM model is suitable for WT in some specific composite fault categories. ADLC, GPC, STMNN, and SVM models aim at detecting a single category while ignoring certain compound categories. The proposed 3DSE-CNN-2DLSTM model considers both single fault categories and some specific compound fault categories, thus enabling the detection of some specific compound categories of WT.

Future research will focus on the following directions: (i) Firstly, to address the problem of data imbalance, this study plans to improve the recognition accuracy of the 3DSE-CNN-2DLSTM model on certain single

fault types by adopting methods such as few-shot learning. Secondly, this study will increase the fault test data and explore different feature engineering methods to capture the features of different fault patterns more accurately. In addition, this study plans to apply transfer learning techniques to transfer the learned fault knowledge to the recognition of other fault types to enhance the generalization ability of the model. These research directions will further enhance the effectiveness and feasibility of the proposed model in practical engineering applications. (ii) This study is based on a data-driven fault detection model; the development of a generic model could be considered for identifying all types of faults, further identifying the type, location, size, and root cause of the fault.

CRediT authorship contribution statement

Tian Wang: Data curation, Formal analysis, Investigation, Software, Validation, Visualization, Writing – original draft. **Linfei Yin:** Conceptualization, Funding acquisition, Project administration, Supervision, Methodology, Resources, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 52107081 and the Guangxi Science and Technology Major Program (China) under Grant AA22068071.

Appendix

Appendix. . Nomenclature

Variables.

- a_i, γ : Hyperparameter.
- b_f, b_i, b_o, b_C : Bias parameter.
- σ_s, σ_g : Sigmoid activation function, hyperbolic tangent activation function
- F_{sg} : Global pooling function
- $f_k, i_k, o_k, \tilde{C}_k, C_k$: Forgetting gate, input gate, output gate, new vector, cell state.
- G : Number of 2D original images.
- g_k : Hidden status updated at time k .
- I, K, U, P : Input channels, output channels, convolution kernel scale, image feature fill width.
- l : The ratio of both descending and ascending dimensions.
- M : Time dimension of the SCADA data.
- μ, ϵ : ReLU function
- N : Feature dimension of SCADA data.
- R_f, R_i, R_o, R_C : Weighting matrix.
- p_i : Predicting the probability of class i th being a positive sample.
- T : Sliding window length.
- V : First layer convolution output as 3D image feature data.
- v_c : Each convolution kernel outputs one channel of output.
- $v_{n,m}, v_c(r,s)$: Element of v_c .
- W, H, C : Width, height, and channels for 3D data.
- W_{conv}, H_{conv} : Width, height after convolution
- W_1, W_2 : Parameters for descending and ascending layers.
- w : Sliding step.

- X:** 2D original image data.
- x_g:** The gth 2D original image.
- x_n:** The nth dimensional vector of gth 2D original image.
- x_{m₀}:** 2D subseries image data after sliding window processing at time m_0 .
- ~X:** 3D image feature data output with the proposed 3DSE-CNN model.
- X_{2D}:** Single-time step 2D image data input with the proposed 2D LSTM model.
- X':** 2D original image data obtained after filling.
- ~x_c:** Output after reweight operation
- ~X:** 3D image feature data.
- x_{n,m}^{U×U}:** Element of input data **X**.
- x_k:** Initial value at moment k .
- y_g:** Label of the gth 2D original image.
- y_{m₀}:** Label of the original 2D image at moment m_0 .
- y:** Image data after squeeze.
- y_c:** The cth element of **y**.
- y':** Image data after excitation operation
- Z:** SCADA data.
- z_m:** State vectors of SCADA system at moment m .
- z_n:** nth dimensional eigenvalue of the state vector at moment t .

References

- Ahakonye, L. A. C., Nwakanma, C. I., Lee, J. M., & Kim, D.-S. (2023). Agnostic CH-DT technique for SCADA network high-dimensional data-aware intrusion detection system. *IEEE Internet of Things Journal*, 1–1. <https://doi.org/10.1109/IoT.2023.3237797>
- Alavi, H., Ohadi, A., & Niaki, S. T. (2022). A novel targeted method of informative frequency band selection based on lagged information for diagnosis of gearbox single and compound faults. *Mechanical Systems and Signal Processing*, 170, Article 108828.
- Attallah, O., Ibrahim, R. A., & Zakzouk, N. E. (2023). CAD system for inter-turn fault diagnosis of offshore wind turbines via multi-CNNs & feature selection. *Renewable Energy*, 203, 870–880.
- Aziz, U., Charbonnier, S., Bérenguer, C., Lebranchu, A., & Prevost, F. (2021). Critical comparison of power-based wind turbine fault-detection methods using a realistic framework for SCADA data simulation. *Renewable and Sustainable Energy Reviews*, 144, Article 110961.
- Cheng, Q. Y., Guang, X. Y., Cheng, M. Y., Yu, Z., & Xi, W. M. (2023). A multi-factor driven spatiotemporal wind power prediction model based on ensemble deep graph attention reinforcement learning networks. *Energy*, 263, Article 126034.
- Chen, W., Qiu, Y., Feng, Y., Li, Y., & Kusiak, A. (2021). Diagnosis of wind turbine faults with transfer learning algorithms. *Renewable Energy*, 163, 2053–2067.
- Dao, P. B. (2022). On wilcoxon rank sum test for condition monitoring and fault detection of wind turbines. *Applied Energy*, 318, Article 119209.
- Deng, W., Li, Z., Li, X., Chen, H., & Zhao, H. (2022). Compound fault diagnosis using optimized MCKD and sparse representation for rolling bearings. *IEEE Transactions on Instrumentation and Measurement*, 71.
- Dhiman, H., Deb, D., Muyeen, S. M., & Kamwa, I. (2021). Wind turbine gearbox anomaly detection based on adaptive threshold and twin support vector machines. *IEEE Transactions on Energy Conversion*, 36(4), 3462–3469.
- Encalada-Davila, A., Moyon, L., Tutiven, C., Puruncajas, B., & Vidal, Y. (2022). Early fault detection in the main bearing of wind turbines based on gated recurrent unit (GRU) neural networks and SCADA data. *IEEE/ASME Transactions on Mechatronics*, 27(6), 5583–5593.
- He, Q., Pang, Y., Jiang, G., & Xie, P. (2021). A spatio-temporal multiscale neural network approach for wind turbine fault diagnosis with imbalanced SCADA data. *IEEE Transactions on Industrial Informatics*, 17(10), 6875–6884.
- Huang, D., Zhang, W. A., Guo, F., Liu, W., & Shi, X. (2023). Wavelet packet decomposition-based multiscale CNN for fault diagnosis of wind turbine gearbox. *IEEE Transactions on Cybernetics*, 53(1), 443–453.
- Kortli, Y., Gabsi, S., Voon, L. F. C. L. Y., Jridi, M., Merzougui, M., & Atri, M. (2022). Deep embedded hybrid CNN-LSTM network for lane detection on NVIDIA jetson xavier NX. *Knowledge-Based Systems*, 240, Article 107941.
- Leahy, K., Hu, R. L., Konstantopoulos, I. C., Spanos, C. J., & Agogino, A. M. (2016). Diagnosing wind turbine faults using machine learning techniques applied to operational data. *2016 IEEE International Conference on Prognostics and Health Management, ICPHM 2016*.
- Leahy, K., Hu, R. L., Konstantopoulos, I. C., Spanos, C. J., Agogino, A. M., & O'Sullivan, D. T. J. (2018). Diagnosing and predicting wind turbine faults from SCADA data using support vector machines. *International Journal of Prognostics and Health Management*, 9(1).
- Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., & Nandi, A. K. (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing*, 138, Article 106587.
- Liao, Y., Huang, W., Shen, C., Zhu, Z., Xuan, J., & Mao, L. (2021). Enhanced sparse regularization based on logarithm penalty and its application to gearbox compound fault diagnosis. *IEEE Transactions on Instrumentation and Measurement*, 70.
- Li, L. L., Liu, Z. F., Tseng, M. L., Jantarakolica, K., & Lim, M. K. (2021). Using enhanced crow search algorithm optimization-extreme learning machine model to forecast short-term wind power. *Expert Systems with Applications*, 184, Article 115579.
- Liu, J., Wang, X., Wu, S., Wan, L., & Xie, F. (2023). Wind turbine fault detection based on deep residual networks. *Expert Systems with Applications*, 213, Article 119102.
- Liu, X., Du, J., & Ye, Z. S. (2022). A condition monitoring and fault isolation system for wind turbine based on SCADA data. *IEEE Transactions on Industrial Informatics*, 18(2), 986–995.
- Liu, Z., Xiao, C., Zhang, T., & Zhang, X. (2020). Research on fault detection for three types of wind turbine subsystems using machine learning. *Energies*, 13(2), 460.
- Li, Y., Liu, S., & Shu, L. (2019). Wind turbine fault diagnosis based on gaussian process classifiers applied to operational data. *Renewable Energy*, 134, 357–366.
- Luo, D., Fang, J., He, H., Lee, W. J., Zhang, Z., Zai, H., ... Zhang, K. (2022). Prediction for dissolved gas in power transformer oil based on TCN and GCN. *IEEE Transactions on Industry Applications*, 58(6), 7818–7826.
- Ma, C., Zhao, Y., Dai, G., Xu, X., & Wong, S. C. (2022). A novel STFSA-CNN-GRU hybrid model for short-term traffic speed prediction. *IEEE Transactions on Intelligent Transportation Systems*.
- Mohine, S., Bansod, B. S., Bhalla, R., & Basra, A. (2022). Acoustic modality based hybrid deep 1D CNN-BiLSTM algorithm for moving vehicle classification. *IEEE Transactions on Intelligent Transportation Systems*.
- Mustafa, Z., Awad, A. S. A., Azzouz, M., & Azab, A. (2023). Fault identification for photovoltaic systems using a multi-output deep learning approach. *Expert Systems with Applications*, 211, Article 118551.
- Pryor, S. C., & Barthelmie, R. J. (2021). A global assessment of extreme wind speeds for wind energy applications. *Nature Energy*, 6(3), 268–276.
- Ritchie, H., Roser, M., & Rosado, P. (2022). Energy. *Our World in Data*. <https://ourworldindata.org/energy>.
- Segu, M., Tonioni, A., & Tombari, F. (2023). Batch normalization embeddings for deep domain generalization. *Pattern Recognition*, 135, Article 109115.
- Shu, X., Yang, J., Yan, R., & Song, Y. (2022). Expansion-squeeze-excitation fusion network for elderly activity recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(8), 5281–5292.
- Subbalakshmi, A., Verma, M., Keerthana, M., Sasimal, S., Harikrishna, P., & Kapuria, S. (2022). Recent advances in experimental and numerical methods for dynamic analysis of floating offshore wind turbines - an integrated review. *Renewable and Sustainable Energy Reviews*, 164, Article 112525.
- Wang, A., Qian, Z., Pei, Y., & Jing, B. (2022). A de-ambiguous condition monitoring scheme for wind turbines using least squares generative adversarial networks. *Renewable Energy*, 185, 267–279.
- Wang, D., Gan, J., Mao, J., Chen, F., & Yu, L. (2023). Forecasting power demand in China with a CNN-LSTM model including multimodal information. *Energy*, 263, Article 126012.
- Wang, Z., Li, G., Yao, L., Qi, X., & Zhang, J. (2021). Data-driven fault diagnosis for wind turbines using modified multiscale fluctuation dispersion entropy and cosine pairwise-constrained supervised manifold mapping. *Knowledge-Based Systems*, 228, Article 107276.
- Ward, T., Johnsen, A., Ng, S., & Chollet, F. (2022). Forecasting SARS-CoV-2 transmission and clinical risk at small spatial scales by the application of machine learning architectures to syndromic surveillance data. *Nature Machine Intelligence*, 4(10), 814–827.
- Wu, Y., & Ma, X. (2022). A hybrid LSTM-KLD approach to condition monitoring of operational wind turbines. *Renewable Energy*, 181, 554–566.
- Xiang, L., Yang, X., Hu, A., Su, H., & Wang, P. (2022). Condition monitoring and anomaly detection of wind turbine based on cascaded and bidirectional deep learning networks. *Applied Energy*, 305, Article 117925.

- Xiao, C., Liu, Z., Zhang, T., & Zhang, X. (2021). Deep learning method for fault detection of wind turbine converter. *Applied Sciences*, 11(3), 1280.
- Xing, S., Lei, Y., Wang, S., Lu, N., & Li, N. (2022). A label description space embedded model for zero-shot intelligent diagnosis of mechanical compound faults. *Mechanical Systems and Signal Processing*, 162, Article 108036.
- Yang, G., Tao, H., Du, R., & Zhong, Y. (2023). Compound fault diagnosis of harmonic drives using deep capsule graph convolutional network. *IEEE Transactions on Industrial Electronics*, 70(4), 4186–4195.
- Yang, J., Wang, L., Song, D., Huang, C., Huang, L., & Wang, J. (2022). Incorporating environmental impacts into zero-point shifting diagnosis of wind turbines yaw angle. *Energy*, 238, Article 121762.
- Yang, Z., Kong, C., Wang, Y., Rong, X., & Wei, L. (2021). Fault diagnosis of mine asynchronous motor based on MEEMD energy entropy and ANN. *Computers & Electrical Engineering*, 92, Article 107070.
- Yan, R., Liao, J., Yang, J., Sun, W., Nong, M., & Li, F. (2021). Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering. *Expert Systems with Applications*, 169, Article 114513.
- Zhang, H., Jiang, L., & Yu, L. (2021). Attribute and instance weighted naive bayes. *Pattern Recognition*, 111, Article 107674.
- Zhang, H., Song, C., Gao, J., Diao, N., & Sun, X. (2022). Image-model-based fault identification for wind turbines using feature engineering and MuSNet. *IEEE Transactions on Industrial Informatics*, 18(10), 6592–6601.
- Zhan, J., Wu, C., Yang, C., Miao, Q., Wang, S., & Ma, X. (2022). Condition monitoring of wind turbines based on spatial-temporal feature aggregation networks. *Renewable Energy*, 200, 751–766.
- Zhu, L., & Song, Y. (2022). Sequential data-driven automatic calibration of wind turbine fault information in smart grids. *IEEE Internet of Things Journal*, 9(11), 8943–8952.
- Zhu, L., & Zhang, X. (2021). Time series data-driven online prognosis of wind turbine faults in presence of SCADA data loss. *IEEE Transactions on Sustainable Energy*, 12(2), 1289–1300.