energies

MDPI

Article

A Novel Attention Temporal Convolutional Network for Transmission Line Fault Diagnosis via Comprehensive Feature Extraction

Guangxun E [1], He Gao [2,3,*], Youfu Lu [1], Xuehan Zheng [2], Xiaoying Ding [3] and Yuanhao Yang [3]

[1] Shandong Hi-Speed Group Co., Ltd., Shandong Hi-Speed Mansion, No. 8 Long'ao North Road, Lixia District, Jinan 250098, China
[2] Shandong Key Laboratory of Intelligent Buildings Technology, School of Information and Electrical Engineering, Shandong Jianzhu University, Jinan 250101, China
[3] Shandong Zhengchen Technology Co., Ltd., No. 11777, Tourist Road, Jinan Area, China (Shandong) Pilot Free Trade Zone, Jinan 250101, China
* Correspondence: gaohe2016@163.com

Abstract: Traditional transmission line fault diagnosis approaches ignore local structure feature information during feature extraction and cannot concentrate more attention on fault samples, which are difficult to diagnose. To figure out these issues, an enhanced feature extraction-based attention temporal convolutional network (EATCN) is developed to diagnose transmission line faults. The proposed EATCN suggests a new comprehensive feature-preserving (CFP) technique to maintain the global and local structure features of original process data during dimension reduction, where the local structure-preserving technique is incorporated into the principal component analysis model. Furthermore, to diagnose transmission line faults more effectively, a CFP-based attention TCN scheme is constructed to classify the global and local structure features of a fault snapshot dataset. To be specific, to cope with the gradient disappearance problem and improve learning capability, a skip connection attention (SCA) network is developed by incorporating a skip connection structure and two fully connected layers into the existing attention mechanism. By combining the developed SCA network with the conventional TCN's residual blocks, an EATCN-based diagnosis model is then constructed to dynamically pay attention to various imported global and local structure features. Detailed experiments on the datasets of the simulated power system are performed to test the effectiveness of the developed EATCN-based fault diagnosis scheme.

Keywords: fault diagnosis; transmission line; comprehensive feature extraction; attention mechanism; temporal convolutional network

check for updates

Citation: E, G.; Gao, H.; Lu, Y.; Zheng, X.; Ding, X.; Yang, Y. A Novel Attention Temporal Convolutional Network for Transmission Line Fault Diagnosis via Comprehensive Feature Extraction. *Energies* **2023**, *16*, 7105. https://doi.org/10.3390/en16207105

Academic Editors: Teke Gush and Raza Haider

Received: 31 August 2023
Revised: 4 October 2023
Accepted: 11 October 2023
Published: 16 October 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Modern power systems are exhibiting rapid development in complexity regarding power transmission and supply in order to satisfy increasing energy requirements [1]. As an important component, transmission lines transmit power from source areas to distribution networks. However, faults frequently occur in transmission lines, impacting the power supply and degrading the reliability of power systems [2,3]. Therefore, to maintain damaged components and reduce downtime, it is of significant importance to accurately diagnose transmission line faults and rapidly eliminate them. Against this background, efficient fault diagnosis schemes for transmission lines are urgently needed to remove these faults and guarantee the safe operation of power systems [4].

With the progress in data measurement and collection systems, massive transmission line running data have become more available. Based on the gathered running data, data-driven-based approaches have a greater ability to diagnose transmission line faults [5,6]. For many data-driven-based transmission line recognition models, original datasets are

often used as model inputs. However, these input datasets are directly collected from the original process variables. Thus, these models cannot cope with the redundant correlations between different original variables [7]. In practice, the variables' redundant correlations lead the process variables to be interrelated and influenced, which always influences the fault recognition effect of the abovementioned methods. Under this circumstance, the performance of these diagnosis approaches for identifying transmission line faults is seriously affected by the variables' redundant correlated properties [8].

Recently, principal component analysis (PCA) has been applied to remove the redundant correlations of original variables and extract the intrinsic latent variables by preserving the main variance information [9,10]. In view of PCA's superiority, a PCA-based approach is applied to characterize the transmission line's running state by retaining the key latent statistics in this paper. However, the traditional PCA-based approach only mines global structure features and neglects important local structure features to eliminate redundant features among the original variables [11]. Local structure features are also significant for feature extraction and dimension reduction because they indicate detailed neighbor relationships between different samples [12]. Missed local structures result in a downside influence on eliminating redundant information and extracting key latent features. Recently, local structure-preserving-based methods have been suggested to excavate the local neighbor structure of samples for feature extraction [13,14] by exploiting the underlying geometrical manifold of the dataset. However, without preserving variance information, the outer shape of the dataset may be broken after the dimension reduction procedure. Hence, the performance of local feature extraction may be degraded if a dataset has significant directions for variance information.

In order to mine the global and local data structures of process data during feature extraction, a novel dimension reduction technique termed comprehensive feature preservation (CFP) is proposed in our work by combining the advantages of PCA and the local structure preservation technique together. In this way, the developed CFP-based feature extraction model exploits the useful global structure information of process data and excavates the original variables' important local data structure features. As a result, the transmission line's fault diagnosis performance is significantly improved by employing the constructed CFP-based feature extraction model to exploit the global and local structure data features during dimension reduction.

After the global and local structure features of the snapshot dataset are extracted, how to effectively analyze these mined fault features is also very significant in recognizing the fault patterns of the transmission line. Recently, deep learning-based approaches have displayed good effectiveness in the fields of fault diagnosis, speech recognition and intelligent transportation due to their superior capability to catch more valuable and useful information from the input data [15–17]. For instance, Guo et al. [18] employed the convolutional neural network (CNN)-based technique to diagnose aircraft sensor faults. Lu et al. [19] adopted modified long short-term memory (LSTM)-based neural networks to identify bearing faults. Based on the deep confidence network, Guo et al. [20] suggested a novel fault recognition approach for the variable refrigerant flow (VRF) system. Furthermore, Zhou [21] discovered that the deep neural network-based approach is more suitable for the VRF system's multiple category fault diagnosis by comparing the five classification approaches. However, most deep learning-based fault diagnosis models, such as CNN and LSTM, contain intricate structures and need numerous preset parameters [22], which can result in the problem of overfitting. As the network layer number increases, the computational complexity increases rapidly, which puts forward much higher demands for hardware devices.

In recent years, convolutional networks have revealed the advantageous ability to deal with the sequence modeling issue. As a typical convolution-based method, temporal convolutional networks (TCNs) have been widely adopted for time-series prediction and pattern recognition [23,24] owing to their simple network structure and abundant neurons. The convolution kernel sharing and parallel computing modules lead the TCN model to

possess reduced computational complexity and shorter computation time. Inspired by the merits of the TCN, a lot of TCN-based approaches have been discussed to enhance the fault diagnosis effectiveness in many applications. For example, Li et al. [7] declared that the TCN-based diagnosis model achieved a much better effect than other neural networks in identifying the chiller fault by managing the process data's coupled dynamics. On account of the TCN's successful applications in the fault diagnosis fields, the fault diagnosis results of the transmission line can be further improved by developing a TCN-based fault recognition method to classify the mined global and local structure features.

The above-introduced TCN-based fault diagnosis algorithms always focus the same attention on different fault samples for the classification instead of affording more attention to the fault samples, which are more difficult to diagnose [25]. These difficultly identified fault samples, indeed, usually have a significant adverse impact on the final fault diagnosis results. Inspired by the emerging transformer network [26], the attention mechanism can be regarded as an effective way to pay more attention to specific fault samples. To more effectively utilize the key features exploited from the original data, the attention mechanism enables fault diagnosis models to possess the capability to pay close attention to different fault samples according to diverse fault patterns, which are widely applied in the fault diagnosis field. For instance, Deng et al. [27] combined the attention mechanism with the LSTM network to set up a sentiment dictionary. Nath et al. [28] employed an amalgamation of a modified attention mechanism-based sensor and transformers to gain an improved performance in rotor fault diagnosis. In addition, based on the feature extraction from time-series images, Fahim et al. [29] integrated another revised attention mechanism into the CNN model to identify the fault pattern of transmission lines. However, the existing attention-based networks always encounter the problem of gradient vanishing for the large number of layers, which results in poor learning ability. Therefore, improvement needs to be carried out to solve this limitation of traditional attention-based networks. After that, if the improved attention mechanism is infused into the existing TCN network for the transmission line fault diagnosis, more attention will concentrate on the fault samples with a significant influence on the fault diagnosis effectiveness, which will further enhance the overall performance of the transmission line's fault identification results.

To exploit much more comprehensive and detailed feature information of the transmission line fault data, an effective combination of the PCA method and the local structure-preserving technique is proposed to mine the data's global and local structure features during the dimension reduction in our work. Furthermore, a novel skip connection attention (SCA) network is established to effectively select and train the parameters of the attention mechanism network. Finally, to achieve much better fault diagnosis effectiveness for the transmission line, the constructed SCA network is integrated into the TCN, referred to as the enhanced attention-based TCN (EATCN), to pay more attention to the extracted global and local structure features of fault samples, which are difficult to recognize and influence the fault identification performance. Based on the above motivations, a comprehensive feature extraction-based EATCN is proposed to identify the transmission line's faults in this paper. The main innovations and contributions are given.

(1) To remove the superfluous correlation information among the process variables, a novel CFP-based dimension reduction model is developed to carry out the feature extraction. The CFP model is capable of mining the global and local structure features of original process data, which are utilized as the input of the developed fault recognition approach. Specifically, considering the traditional PCA only maintains the global structures feature for the dimension reduction, the local structure-preserving technique is combined into the PCA to establish the CFP model;

(2) To train the network's parameters more effectively and to improve the network's learning capability, the skip connection attention (SCA) network is proposed. To be specific, to establish the issue of gradient vanishing for the masses of layers when training the parameters, the skip connection is incorporated into the conventional attention mechanism

to construct the SCA network. In addition, the two fully connected layers are adopted in the SCA to enhance the network's generalization ability;

(3) To further reinforce the transmission line's fault recognition performance, the established SCA network is integrated into the TCN model to classify the global and local structure features extracted by the CFP. By combing the SCA network with the TCN's residual blocks, the suggested EATCN is set up to derive and stack up the modified residual blocks. The fault diagnosis model's input, i.e., the mined global and local structure features, is recoded by the developed EATCN, which guarantees that more attention is focused on the valuable information and internal correlations of the imported features;

(4) To prove the fault diagnosis effect of the proposed EATCN-based transmission line diagnosis scheme, the EATCN model is compared with some closely related fault diagnosis models. The experiments of suggested EATCN and comparisons with related approaches are carried out on the datasets of the simulated power system. In contrast to the related diagnosis models, the fault diagnosis results display the superior effectiveness of the EATCN-based recognition scheme.

2. Preliminaries

2.1. The Short Circuit Faults of Transmission Line

The transmission line usually incurs short circuit faults, which always lead to residual life reduction, power loss increase and so on [30]. Different short circuit faults always arise during the transmission line's daily running, which can be divided into asymmetrical faults and symmetrical faults [3]. To be specific, the line-to-ground (LG) faults, the line-to-line (LL) faults and the double lines-to-ground (LLG) faults are asymmetrical faults, which possess much bigger occurrence probability compared with symmetrical faults [31]. On the other hand, the triple line (LLL) faults and triple lines-to-ground (LLLG) faults remain symmetrical faults, which result in much greater damage to the transmission line than asymmetrical faults [31]. However, the occurrence probability of these symmetrical faults is smaller than that of the asymmetrical faults.

Short circuit faults seriously influence the running status of the transmission line; therefore, the rapid and accurate recognition of the short circuit faults is important to ensure the safe and stable operation of the transmission line [32,33]. According to the specific fault pattern, the amplitude, phase and frequency of voltages and currents for the short circuit faults would incur obvious changes, and these changes are unique for the different fault patterns. In this way, the covariate features can be captured from the short circuit faults' temporal operation data to discern different fault patterns as well as process disturbances.

2.2. The Basic PCA Method

The PCA possesses the ability to cope with highly redundant process data through mapping the redundant data into a low-dimensional principal component subspace by containing the most variations of the data [9–11]. Therefore, the PCA is widely used for data feature extraction.

Suppose $\boldsymbol{X} \in \boldsymbol{R}^{n \times m}$ is the original high-dimensional dataset, which contains n samples, and each sample consists of m measured variables. After the dataset $\boldsymbol{X}$ is scaled according to the normal operation dataset, the covariance $\boldsymbol{S}$ of the dataset $\boldsymbol{X}$ is first computed.

$$\boldsymbol{S} = \frac{1}{n-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} \tag{1}$$

Then, Equation (2) is carried out on the obtained covariance, which is given as follows:

$$\boldsymbol{S} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^{\mathrm{T}} \tag{2}$$

where $\boldsymbol{D}$ represents a diagonal matrix, whose diagonal contains the decreasing order eigenvalues $\widetilde{\lambda}_1 > \widetilde{\lambda}_2 > \ldots > \widetilde{\lambda}_{rank(\boldsymbol{X})}$ of the matrix $\boldsymbol{S}$. The l eigenvectors $\widetilde{\boldsymbol{p}}_1, \widetilde{\boldsymbol{p}}_2, \ldots, \widetilde{\boldsymbol{p}}_l$ related to the first l largest eigenvalues $\widetilde{\lambda}_1 > \widetilde{\lambda}_2 > \ldots > \widetilde{\lambda}_l$ are retained to construct the loading matrix $\widetilde{\boldsymbol{P}} = [\widetilde{\boldsymbol{p}}_1, \widetilde{\boldsymbol{p}}_2, \ldots, \widetilde{\boldsymbol{p}}_l] \in R^{m \times l}$.

Based on the loading matrix $\widetilde{\boldsymbol{P}}$, the original dataset $\boldsymbol{X}$ is mapped into the PC and residual spaces.

$$\boldsymbol{X} = \boldsymbol{T}\widetilde{\boldsymbol{P}}^{\mathrm{T}} + \boldsymbol{E} = \boldsymbol{t_1}\widetilde{\boldsymbol{p}_1}^{\mathrm{T}} + \boldsymbol{t_2}\widetilde{\boldsymbol{p}_1}^{\mathrm{T}} + \cdots + \boldsymbol{t_l}\widetilde{\boldsymbol{p}_l}^{\mathrm{T}} + \boldsymbol{E} \tag{3}$$

where $\boldsymbol{T}$ indicates the score matrix $\boldsymbol{T} = [\boldsymbol{t_1}, \boldsymbol{t_2}, \cdots, \boldsymbol{t_l}]$ and $\boldsymbol{t_i} \in R^n$ is the i-th score vector. Note that the vector $\widetilde{\boldsymbol{p}_i}$ is also called the i-th loading vector and l also represents the principal component (PC) number retained in the principal space.

The residual matrix $\boldsymbol{E}$ in residual space is computed as

$$\boldsymbol{E} = \boldsymbol{X} - \hat{\boldsymbol{X}} = \boldsymbol{X} - \boldsymbol{T}\widetilde{\boldsymbol{P}}^{\mathrm{T}} \tag{4}$$

In the PC space, the computed PCs can represent the main changes in the process, while the latent variables in the residual space characterize the process noise information.

2.3. The Basic Temporal Convolutional Network

For the time-series modeling tasks, the TCN-based methods have revealed superior effectiveness than the LSTM, GRU and their improved approaches [23] due to the TCN's simple and clear network structure without the complicated gate computing units. Furthermore, similar to CNN, another merit of TCN is that it applies parallel computation and shares the convolution kernel to cut down the computational complexity [24].

Compared with the LSTM containing the complicated gate computing units, the TCN has the superiority to sufficiently excavate the dynamic property among the process variables because of the TCN's unique causal convolution layer displayed in Figure 1. This unique convolution layer not only guarantees that the current output of the TCN is derived based on only the current and previous input samples with low computational complexity but can also cope with the TCN's gradient vanishing issue [23]. The receptive field is bigger, and the convolution layers' number is smaller [24]. More precisely, the convolution computation of the input vector's n-th element is given as follows.

$$F(n) = (\boldsymbol{x}_t * dq)(n) = \sum_{r=0}^{k-1} q(r) \cdot \boldsymbol{x}_{t,\, n-d\cdot r} \tag{5}$$

where $\boldsymbol{x}_t$ denotes the input vector, q represents the filter with a size of k and d indicates the dilation coefficient.

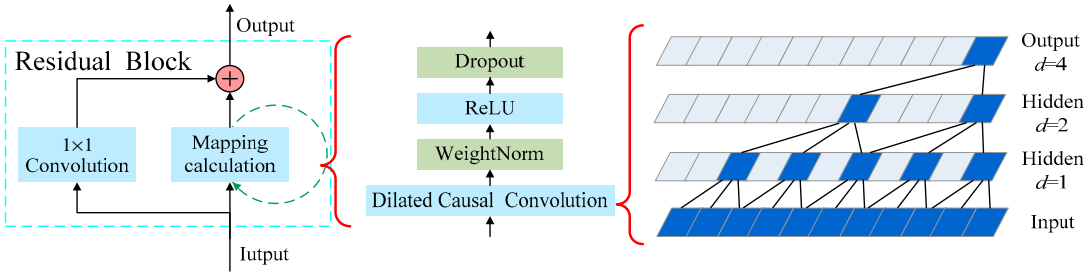


Figure 1. A schematic diagram of the basic TCN network.

In addition, the residual block illustrated in Figure 1 is adopted in the TCN model to strengthen the classification performance by settling the troublesome network degradation. To obtain an equal width for the output and input, the residual block's input is imported into a 1×1 convolution layer. Meanwhile, the input vector is also fed to the mapping calculation to establish the final output. As an important component of the mapping calculation, the rectified linear unit (ReLU) denotes a nonlinear activation function, which is utilized to determine the maximum between zero and the value of the input. With the help of the ReLU, the TCN can fit any function, which contributes to achieving fault diagnosis effectiveness. In summary, the TCN model is established to stack the residual blocks, which are constructed by carrying out the dilated and causal convolution operations.

3. The Developed CFP-Based Feature Extraction Technique

As introduced in Section 2, the PCA cannot grasp the local structure features of the transmission line data. Motivated by this, to further remove the redundant correlation information of the original process variables, a novel comprehensive feature-preserving (CFP)-based dimension reduction technology is proposed to mine the global and local structure features of the original process data, which incorporate the locality structure-preserving technique into the PCA model.

Let the matrix $\boldsymbol{X} = [\boldsymbol{x}(1), \boldsymbol{x}(2), \cdots, \boldsymbol{x}(n)]^T \in \boldsymbol{R}^{n \times m}$ with n normal samples and m variables denote the normal operating dataset. The training matrix **X** is first scaled by subtracting its mean and dividing its standard deviation. Based on the normalized matrix **X**, the PCA model seeks a loading vector **p,** which can guarantee that the distance among all the samples in the PC space is maximized.

$$J(\boldsymbol{p})_{PCA} = \max \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{p}^T (\boldsymbol{x}(i) - \bar{\boldsymbol{x}})(\boldsymbol{x}(i) - \bar{\boldsymbol{x}})^T \boldsymbol{p} = \max \boldsymbol{p}^T \boldsymbol{G} \boldsymbol{p} \tag{6}$$

where $\bar{\boldsymbol{x}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}(i)$ is the mean value of n samples, and $\boldsymbol{G} = \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{x}(i) - \bar{\boldsymbol{x}})(\boldsymbol{x}(i) - \bar{\boldsymbol{x}})^T$.

From Equation (6), it can be found that the PCA only preserves the global structure features while clearing off the redundant correlations. To optimally preserve the global and local structure features of the training matrix **X**, the local structure-preserving framework is combined with the PCA model in our work.

For the i-th sample $\boldsymbol{x}(i)$, its nearest neighbors $\boldsymbol{x}(j)$, $j = 1, 2, \ldots, K$ are first searched to construct the local neighborhood subset $\boldsymbol{X}_{\text{sub}}(i)$ by means of the k-nearest neighbor approach. Therefore, the acquired neighbor matrix $\boldsymbol{X}_{\text{sub}}(i)$ has the ability to reveal if the j-th sample $\boldsymbol{x}(j)$ belongs to the local neighborhood of the $\boldsymbol{x}(i)$. To be specific, the element $\boldsymbol{W}_{ij}$ of the similarity matrix $\boldsymbol{W}$ is determined as:

$$\boldsymbol{W}_{ij} = \begin{cases} w\{\boldsymbol{x}(i), \boldsymbol{x}(j)\} = \exp(-\|\boldsymbol{x}(i) - \boldsymbol{x}(j)\|^2), & if\ \boldsymbol{x}(j) \in \boldsymbol{X}_{\text{sub}}(i)\ or\ \boldsymbol{x}(i) \in \boldsymbol{X}_{\text{sub}}(j) \\ 0, & otherwise \end{cases} \tag{7}$$

where $w\{\boldsymbol{x}(i), \boldsymbol{x}(j)\}$ indicates the neighborhood relationship between the samples $\boldsymbol{x}(i)$ and $\boldsymbol{x}(j)$; thus, the matrix $\boldsymbol{W}$ can represent the local neighbor relations of the training matrix $\boldsymbol{X}$.

The local structure-preserving (LSP) framework aims for the loading vector $\boldsymbol{p}$ to hold the local neighbor relations of the training matrix $\boldsymbol{X}$ by minimizing the distances of neighbor samples in the PC space.

$$\begin{aligned} J(\boldsymbol{p})_{LSP} &= \min \tfrac{1}{2} \textstyle\sum_{i=1}^{n} \boldsymbol{p}^T (\boldsymbol{x}(i) - \boldsymbol{x}(j)) \boldsymbol{W}_{ij} (\boldsymbol{x}(i) - \boldsymbol{x}(j))^T \boldsymbol{p} \\ &= \min \boldsymbol{p}^T \boldsymbol{X} (\boldsymbol{D} - \boldsymbol{W}) \boldsymbol{X}^T \boldsymbol{p} = \min \boldsymbol{p}^T \boldsymbol{X} \boldsymbol{L} \boldsymbol{X}^T \boldsymbol{p} \end{aligned} \tag{8}$$

where $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$ represents the Laplacian matrix [12–14], and **D** is a diagonal matrix with the i-th element as $\boldsymbol{D}_{ii} = \sum_{j=1}^{n_s} \boldsymbol{W}_{ji}$, $i = 1, 2, \cdots, n_s$.

To exploit the local and global structure features of the training matrix during the dimension reduction, the optimization $J(\boldsymbol{p})_{CFP}$ of the proposed CFP technique is constructed in Equation (9) to derive the optimal loading vector $\boldsymbol{p}$, which simultaneously maximizes the PCA's objective function and minimizes the optimization of the local structure-preserving framework.

$$\begin{aligned} J(\boldsymbol{p})_{CFP} &= \max \{\eta J(\boldsymbol{p})_{PCA} - (1 - \eta) J(\boldsymbol{p})_{LSP}\} \\ &= \max \boldsymbol{p}^T \{\eta \boldsymbol{G} - (1 - \eta) \boldsymbol{X} \boldsymbol{L} \boldsymbol{X}^T\} \boldsymbol{p} = \max_{a} \boldsymbol{p}^T \boldsymbol{Q} \boldsymbol{p} \\ s.t.\ \ & \boldsymbol{p}^T \boldsymbol{p} = 1 \end{aligned} \tag{9}$$

where $J(\boldsymbol{p})_{PCA}$ is utilized to keep the global structure features of the training matrix and the local structure features are retained by $J(\boldsymbol{p})_{LSP}$. The η is a tradeoff parameter to balance the optimizations $J(\boldsymbol{p})_{PCA}$ and $J(\boldsymbol{p})_{LSP}$. The matrix $\boldsymbol{Q}$ is calculated as:

$$\boldsymbol{Q} = \eta \boldsymbol{G} - (1 - \eta) \boldsymbol{X} \boldsymbol{L} \boldsymbol{X}^T \tag{10}$$

Equation (9) can be solved by computing the eigenvalue decomposition defined in Equation (11).

$$\boldsymbol{Qp} = \lambda \boldsymbol{p} \tag{11}$$

Suppose $\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_d$ are the eigenvectors of the first d largest eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_d$. Then, the loading matrix $\boldsymbol{P}$ of the CFP is built by retaining these d eigenvectors $\boldsymbol{P} = [\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_d] \in \boldsymbol{R}^{n \times d}$. These loading vectors are mutually orthogonal, which can effectively improve the discriminative ability of the CFP-based dimension reduction method when extracting the global and local structure features of original data.

The number of the retained loading vectors d in the PC space is determined based on the cumulative contribution rate, which is given as follows:

$$R_d = \sum_{i=1}^{d} \lambda_i / \sum_{j=1}^{m} \lambda_j \tag{12}$$

In this paper, the value of d in the CFP model is selected by the 95% cumulative contribution rate.

Based on the loading matrix $\boldsymbol{P}$, the training matrix $\boldsymbol{X}$ is decomposed by the suggested CFP model.

$$\boldsymbol{X} = \boldsymbol{YP}^T + \tilde{\boldsymbol{X}} \tag{13}$$

where $\boldsymbol{Y} = \boldsymbol{XP}$ is the score matrix, which is composed of the derived global and local structure features in the feature space and $\tilde{\boldsymbol{X}} = \boldsymbol{X} - \boldsymbol{YP}^T$ indicates the residual matrix.

When the fault sample $\boldsymbol{x}_F$ becomes available, the latent significant features $\boldsymbol{y}_F$ of the $\boldsymbol{x}_F$ are then extracted by protecting $\boldsymbol{x}_F$ in the feature space, which is expressed as follows.

$$\boldsymbol{y}_F = \boldsymbol{P}^T \boldsymbol{x}_F \tag{14}$$

After the snapshot dataset $\boldsymbol{X}_F = [\boldsymbol{x}_F(1), \boldsymbol{x}_F(2), \ldots, \boldsymbol{x}_F(n_f)]^T$ is set up based on the gathered n fault samples, the dataset $\boldsymbol{X}_F$ is then normalized using the training matrix $\boldsymbol{X}$. Thereafter, the built CFP is applied to extract the global and local structure features of the snapshot dataset $\boldsymbol{X}_F$ and the historical fault datasets. To gain improved fault diagnosis effectiveness, these exploited global and local structure features are regarded as the input of the subsequently developed recognition model.

4. The Enhanced Attention TCN-Based Fault Diagnosis Model

To strengthen the fault recognition effect of the transmission line, a new SCA network is set up, and the enhanced attention TCN (EATCN)-based diagnosis model is established by integrating the suggested SCA into the conventional TCN. To be specific, to dispose of the issue of gradient vanishing and improve the learning ability for a large number of network layers, a skip connection structure and two fully connected layers are infused into the conventional attention mechanism to build the SCA network. The proposed SCA is then applied to modify the residual blocks of the existing TCN to provide dynamic attention to the inputted global and local structure features, which intensifies the transmission line's fault identification effect because the difficultly diagnosed features are focused on more attention. Below, we will describe the details of the suggested SCA mechanism and the constructed SCA-based EATCN diagnosis model.

4.1. The Established SCA Network

The attention mechanism is skilled in catching the data's internal relations by mapping the query and key values to the network's output [26]. In comparison with the existing attention mechanism, the developed SCA network employs a structure similar to the mobilenet V3 [34] and skip connection structure [35]. The SCA has the ability to recode the imported global and local structure features, which makes the ATCN-based recognition model pay close attention to the valuable information of the input.

A schematic diagram of the proposed SCA network is illustrated in Figure 2. The three intermediate vectors **value**, **query** and **key** are derived from the imported global and

local structure features extracted by the CFP by means of the three respective input layers. As exhibited in Figure 2, the input layers contain two fully connected layers to exploit the important interrelations between the input features and the three intermediate vectors. In addition, the fully connected layers are able to ensure that the input and output have the same dimension. That is, the dimension of the input global and local structure features is equal to that of the three intermediate vectors. Specifically, based on the matrices $\boldsymbol{V}$, $\boldsymbol{Q}$ and $\boldsymbol{K}$ of the three intermediate vectors, the output of the constructed SCA network is formulated as

$$output = F_{score}(\boldsymbol{Q}, \boldsymbol{K}) \cdot \boldsymbol{V} \tag{15}$$

where the $\boldsymbol{F}_{score}$ denotes the attention scores calculated by the softmax layer in the SCA network.

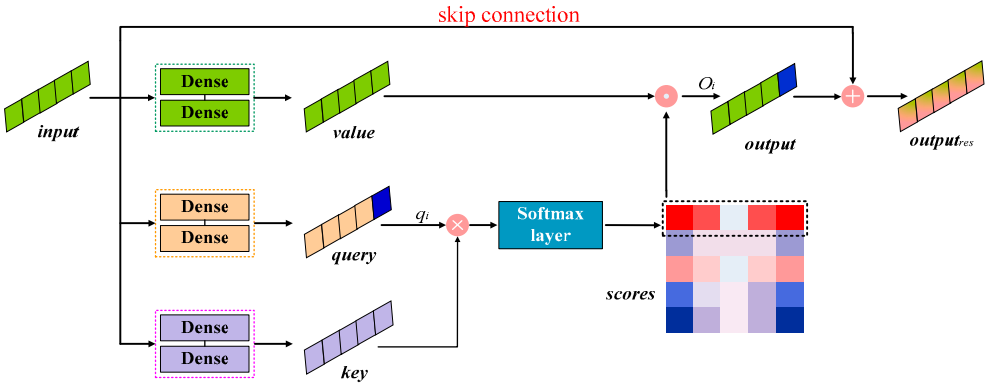


Figure 2. The structure of the developed SCA network.

With the help of the SCA's softmax layer, the score vector is derived by multiplying the vector **query**'s element q_i with the elements in the vector **key**. In this way, the score matrix can be achieved by repeating the above multiplication operation. Then, the vector $\boldsymbol{c}$'s element o_i is worked out by multiplying the vector **score** with the vector **value**. Note that the vector **score**'s elements denote the attention paid to the homologous elements in the vector **value**, and the summation of these elements is equal to one.

Furthermore, the skip connection structure is incorporated into the existing attention network to work out the gradient vanishing issue. Based on the computed vector **output**, the residual output vector $\mathbf{output}_{\mathrm{res}}$ is further derived by superposing the vector **input** on the vector **output**, i.e., $\boldsymbol{output}_{res} = \boldsymbol{input} + \boldsymbol{output}$. The added skip connection can ensure the vector's **input** and $\mathbf{output}_{\mathrm{res}}$ have the same data dimension by means of indirectly merging the data in channels.

4.2. The Developed EATCN Fault Diagnosis Model Based on the SCA Network

In order to utilize the different attention to the imported global and local structure features, i.e., focus more attention on these difficultly diagnosed features, the EATCN model is further set up by combining the developed SCA network with the residual blocks. Figure 3 illustrates a flowchart of the EATCN's residual block to avert the gradient vanishing, which is inspired by the residual connection from the residual neural network. As exhibited in Figure 4, the EATCN also adopts the dilated convolution network to further cut down the model parameters while offering enough of a receptive field. Given an input vector $\mathbf{x}$ with n elements, the dilated convolution operation is expressed as

$$F(n) = (\boldsymbol{x} * d\boldsymbol{l})(n) = \sum_{r=0}^{f-1} \boldsymbol{l}(r) \cdot \boldsymbol{x}_{n-d \cdot r} \tag{16}$$

where $\boldsymbol{l} \in (0, 1, \ldots, k-1)$ and f, respectively, represent the filter and its size. The adjacent nodes' interval changes along with the value of the dilation factor d varying. In addition, enough key features can be exploited at multiple scales in the EATCN model because the different values of the dilation factor in the multiple layers result in various receptive fields.

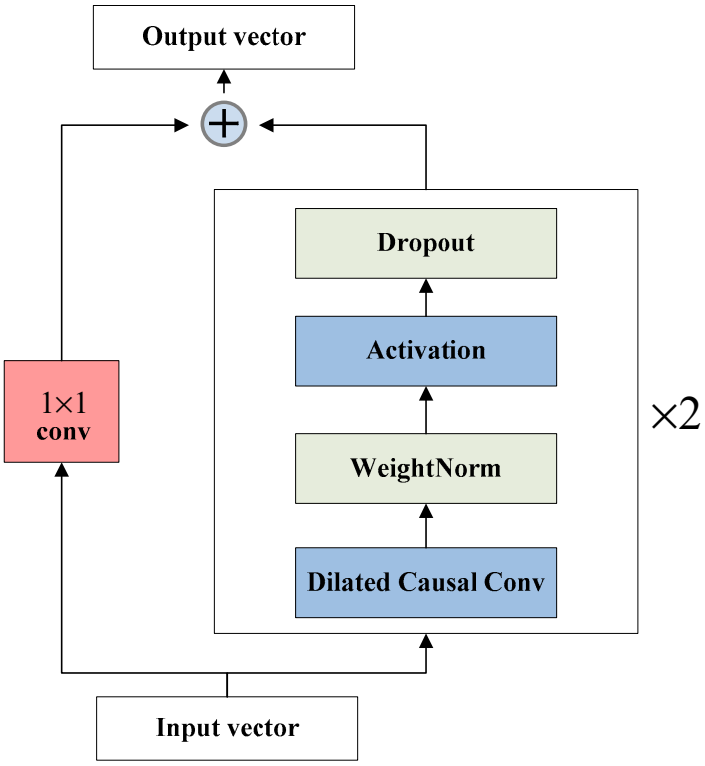


Figure 3. The flowchart of the EATCN's residual block.

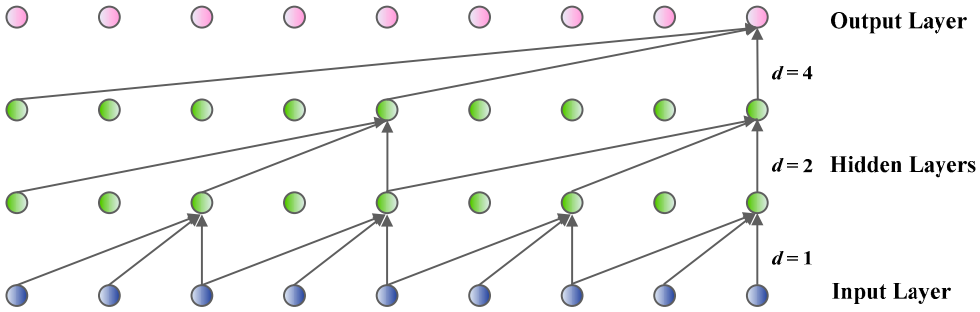


Figure 4. The structure chart of dilated convolution in the EATCN.

A schematic diagram of the SCA network-based EATCN fault diagnosis model is revealed in Figure 5. It is observed that the constructed SCA network is placed at the front of the EATCN's residual block to constitute the improved residual block, and the EATCM model is set up by stacking these improved residual blocks. With the help of the infused SCA network, the EATCN-based fault diagnosis model possesses the capability to dynamically pay different levels of attention to the input global and local structure features, which can effectively identify the fault pattern of the not easily distinguishable features. Due to the SCA network, the output of the previous improved residual block is recoded by the current improved residual block, which makes each improved residual block pay more attention to the more valuable features mined by the CFP-based dimension reduction model.

According to the above analysis and formulation, the SCA network-based EATCN fault diagnosis model has the superiority to pay more attention to the imported global and local structure features, which are difficult to identify. Therefore, by adding the dense and softmax layers into the EATCA model, the overall fault recognition effectiveness can be further heightened by employing the suggested EATCN as a powerful classifier to analyze the transmission line's running data and discern the fault pattern.

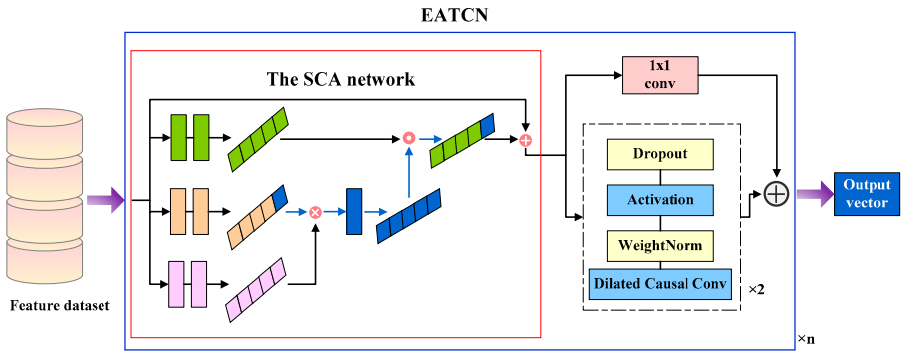


Figure 5. The SCA network-based EATCN fault diagnosis model.

5. The EATCN-Based Fault Diagnosis Scheme for the Transmission Line

Figure 6 displays the enhanced feature extraction-based attention TCN (EATCN) fault diagnosis strategy for the transmission line. As shown in Figure 6, in the modeling stage, based on the normal operating dataset $\boldsymbol{X}$, the developed CFP model is first established by infusing the local structure-preserving in the PCA model, and then the built CFP model is employed to extract the global and local structure features of the C class historical fault datasets. The skip connection structure and two fully connected layers are subsequently merged into the existing attention mechanism to establish a new SCA network, and the EATCN-based diagnosis model is set up through incorporating the suggested SCA network with each residual block in the conventional TCN model. Finally, the mined historical global and local structure features are imported into the EATCN network to train the fault diagnosis model. In the fault diagnosis stage, the constructed CFP is first adopted to exploit the global and local structure features of the fault snapshot dataset $\boldsymbol{X}_F$, and the trained EATCN is applied to classify the extracted global and local structure features to recognize the pattern of the snapshot dataset $\boldsymbol{X}_F$. Due to EATCN's virtue of paying more attention to the exploited structure features, which are difficult to classify, the pattern of the snapshot dataset can be effectively and accurately identified.

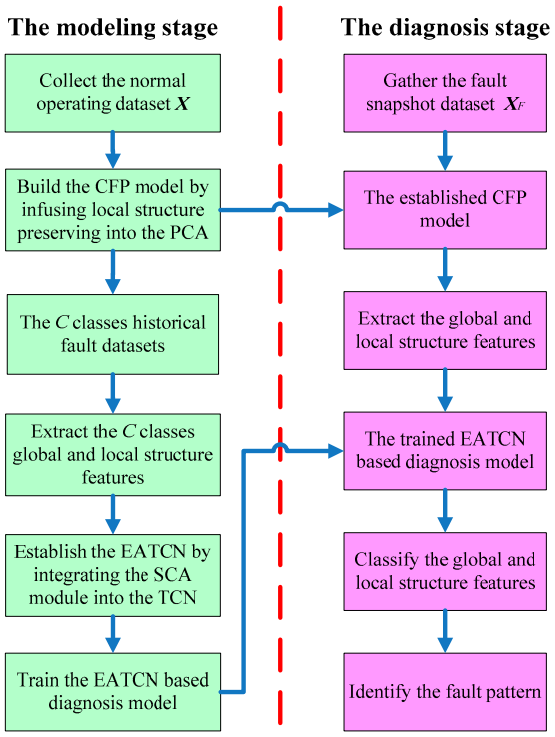


Figure 6. The enhanced feature extraction-based EATCN fault diagnosis strategy.

6. The Experiments and Comparisons

6.1. Introduction of the Experimental Data

A benchmark power system is modeled in the MATLAB/Simulink environment [3,36,37] to simulate the normal operating and multiple fault datasets, whereas Simscape Electrical affords the component library to model electronic, mechatronic and electrical power systems. The simulated system is widely applied to perform power system studies, which is discussed in the literature [3,37]. As displayed in Figure 7, the simulated power system includes two symmetrical areas, which are connected by the 220 km length transmission line. Except for the transmission line, the faults may occur on other components of the simulated power system, such as transformers, generators, incipient short circuit faults at underground cables and so on. However, because our work focuses on identifying patterns in the short circuit faults on transmission lines, the eleven fault patterns listed in Table 1 are simulated and studied at the 220 km length transmission line of the simulated power system, i.e., the region from V_7 to V_8 and the region from V_8 to V_9.

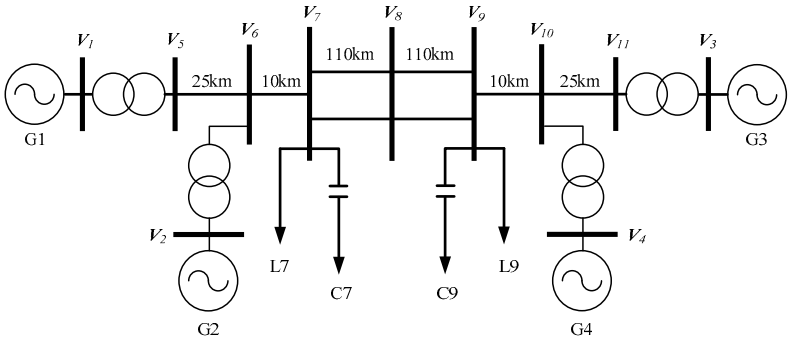


Figure 7. Single line diagram of the simulated benchmark power system [3,36,37].

Table 1. The description of normal running and eleven fault patterns.

Number	Fault Pattern	Fault Description
0	NF	No fault (normal operation)
1	AG	Short fault of line A to ground
2	BG	Short fault of line B to ground
3	CG	Short fault of line C to ground
4	AB	Short fault of line A to line B
5	BC	Short fault of line B to line C
6	AC	Short fault of line A to line C
7	ABG	Short fault of line A and line B to ground
8	BCG	Short fault of line B and line C to ground
9	ACG	Short fault of line A and line C to ground
10	ABC	Short fault of line A, line B and Line C
11	ABCG	Short fault of line A, line B and Line C to ground

The power system is simulated under normal running and short circuit fault conditions on the transmission line. Comprehensive normal and fault datasets are constructed to build the developed EATCN-based fault diagnosis model by measuring and collecting the line voltages and currents of the power system. A total of 12,000 samples are generated and labeled for twelve types of operating condition, which includes one normal operating condition and eleven short circuit fault patterns. Thus, each type of operating condition contains 1000 samples. Before the normalization of transmission line datasets, the Gaussian noise with zero mean and 0.01 variance is introduced to the monitored variables for the purpose of simulating the actual measurement noise. As listed in Table 1, the simulated eleven fault patterns are {AG, BG, CG, AB, BC, AC, ABG, BCG, ACG, ABC, ABCG}, where the symbols A, B, C and G, respectively, stand for the phases A, B, C and ground. These eleven short circuit fault patterns are classified as double line (LL) faults, line-to-ground

(LG) faults, triple line (LLL) faults, double line-to-ground (LLG) faults and triple line-to-ground (LLLG) faults, where only the LLL and LLLG fault patterns are symmetric faults and the remaining are asymmetric faults.

To diagnose the fault pattern of the transmission line, the first 400 collected fault samples are utilized to build the snapshot dataset. The remaining 600 fault samples from the same pattern are regarded as the historical fault dataset. The EATCN-based fault diagnosis model is first trained by feeding the mined historical fault dataset's global and local structure features, which are extracted by the developed CFP approach, and the snapshot dataset's global and local structure features are then extracted and imported to the established EATCN-based diagnosis model for the purpose of identifying the pattern of detected faults.

6.2. Compared Approaches and Effectiveness Evaluation Index

To train and deploy the proposed EATCN, the experiments are conducted in the MATLAB R2020a computational environment, which runs on a computer with an Intel(R) Core (TM) i7-10750H CPU @ 2.60GHz and 16.0 GB (15.7 GB usable) installed RAM. In addition, to prove the effect of the proposed EATCN-based diagnosis approach, some traditional fault diagnosis methods, i.e., the support vector machine (SVM), the deep belief network (DBN) and the long short-term memory (LSTM) network, are contrasted with the suggested EATCN. The global and local structure features derived by the constructed CFP are imported to the SVM, DBN and LSTM.

To train the EATCN-based diagnosis model, a tradeoff parameter is chosen as 0.5 by experience; the threshold of the cumulative contribution rate is determined as 95% through trial and error. The learning rate is chosen as 0.001, the batch size is determined as 64, the number of hidden units is 500, and the expansion factor is 2, according to the cross-validation. During the LSTM training, the number of hidden units is set to 300, the batch size is 64 and the 0.001 learning rate is utilized through trial and error. In addition, the optimal model parameters are determined using the Adam optimizer. The node numbers of the DBN's three convolution layers are, respectively, selected as 32, 64 and 128 through trial and error. The batch size and the learning rate in the DBN are also, respectively, chosen as 64 and 0.001 for fairness. In the SVM, the Gaussian kernel function is adopted, and the parameter is set to 600 using the grid search method. In addition, the weight factor of the SVM is experientially determined as 50.

To assess the performance of the discussed EATCN for the transmission line fault diagnosis, four performance indices, i.e., the fault diagnosis rate $FDR(i),\ i = 1, 2, \ldots, C$ of the fault samples in the i-th pattern, the average fault diagnosis rate FDR_{average} of fault samples in the total of C patterns, the precision $P(i), i = 1, 2, \ldots, C$ for the i-th pattern and the average precision P_{average} for a total of C patterns, are utilized.

Particularly, the index $FDR(i)$ is defined as

$$FDR(i) = \frac{N_i(i)}{N_i},\ \ i = 1, \cdots, C \tag{17}$$

where $N_i(i)$ denotes the number of correctly diagnosed fault samples for the i-th pattern.

The index FDR_{average} expressed in Equation (18) indicates the average value of all the acquired fault diagnosis rates for the C fault snapshot datasets.

$$FDR_{\text{average}} = \frac{1}{C}\sum_{i=1}^{C} FDR(i) \tag{18}$$

The precision $P(i),\ i = 1, 2, \ldots, C$ for the i-th pattern is defined as:

$$P(i) = \frac{N_i(i)}{(N_i(i) + FP_i(i))} \tag{19}$$

where $FP_i(i)$ represents the number of wrongly identified fault samples in the i-th pattern.

The index average precision P_{average} expressed in Equation (20) indicates the average value of all the computed precisions for the C fault snapshot datasets.

$$P_{\text{average}} = \frac{1}{C}\sum_{i=1}^{C} P(i) \tag{20}$$

6.3. Comparison of the Fault Diagnosis Results

(1) Fault diagnosis results comparison for the pattern ABC

After the snapshot dataset, $\boldsymbol{S}_{\text{ABC}}$ of the short circuit fault ABC is gathered, the index *FDR*'s values of SVM, DBN, LSTM and EATCN for the dataset $\boldsymbol{S}_{\text{ACG}}$ are computed as 64.25%, 71.75%, 79.00% and 93.50%. It is observed that the SVM displays the worst fault identification effect, while the DBN and LSTM have better index values for *FDR*, i.e., 71.75% and 79.00%, which still needs to be enhanced. Different from SVM, DBN and LSTM, EATCN gains the best recognition capability for the dataset $\boldsymbol{S}_{\text{ABC}}$ with the largest value of the index *FDR*, i.e., 93.50%. This is due to EATCN's outstanding performance in exploiting and classifying the global and local structure features contained in the running data of the transmission line. Again, the histogram of the fault identification results for these four fault diagnosis models is illustrated in Figure 8, which clearly reveals that the diagnosis effectiveness of the discussed EATCN is much better than that of the SVM, DBN and LSTM for discerning the fault pattern ABC.

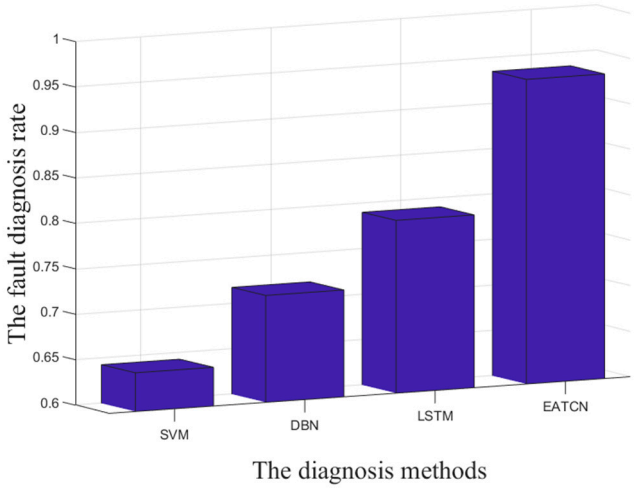


Figure 8. The diagnosis results of the SVM, DBN, LSTM and EATCN for the fault ABC.

(2) Fault diagnosis results comparison for the pattern BC

The values of the index *FDR* for the SVM, DBN, LSTM and EATCN for the short circuit fault BC turn out to be, respectively, 76.50%, 70.75%, 81.75% and 92.50%. It is obvious that there are different fault recognition results for these four diagnosis models to discern the dataset $\boldsymbol{S}_{\text{BC}}$. Compared with the LSTM and EATCN, the DBN and SVM exhibit a much more unsatisfied fault identification effect to diagnose the fault pattern BC. On the contrary, LSTM reveals the improved diagnosis effect because LSTM's index *FDR* value is 81.75%. In the end, the EATCN reveals the best fault recognition effect as its *FDR* value is 92.50%. To make a more vivid comparison, the values of the index *FDR* for the SVM, DBN, LSTM and EATCN are plotted in the histogram in Figure 9. According to the above analysis, the advantage of the EATCN is fully certified over the LSTM, DBN and SVM when identifying the pattern BC.

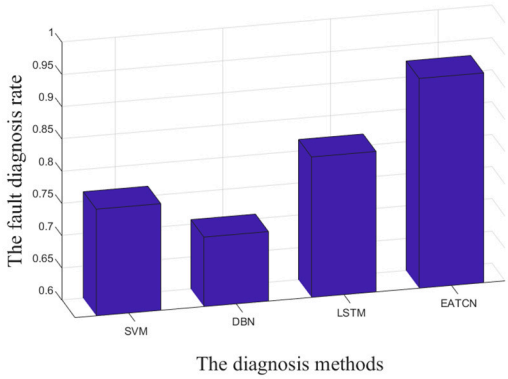


Figure 9. The diagnosis results of the SVM, DBN, LSTM and EATCN for the fault BC.

(3) Fault diagnosis results comparison for the pattern ABG

After the fault snapshot dataset S_{ABG} is constructed, the SVM-, DBN-, LSTM- and EATCN-based fault diagnosis models are employed to identify the pattern of short circuit fault ABG by computing the performance index *FDR*. To be specific, the values of the index *FDR* for the SVM, DBN, LSTM and EATCN are, respectively, 71.50%, 83.50%, 87.25% and 94.25%, which proves that the EATCN possesses the highest value of the index *FDR* among these four fault diagnosis models. To make a more visualized analysis, the four different *FDR* values are further represented by a bar chart in Figure 10. In this way, it can be concluded that the EATCN-based diagnosis model outperforms the LSTM, DBN and SVM in terms of recognizing the fault pattern ABG.

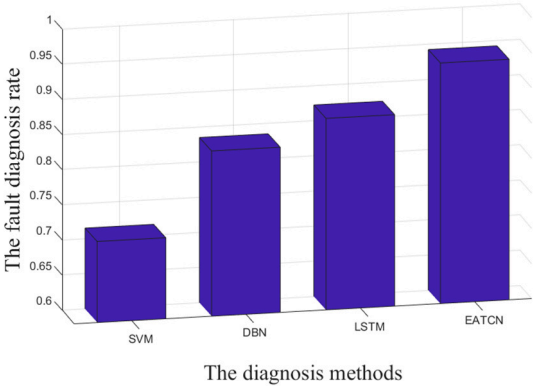


Figure 10. The diagnosis results of the SVM, DBN, LSTM and EATCN for the fault ABG.

(4) Fault diagnosis results comparison for the eleven patterns

After the above-introduced eleven fault patterns are detected, fault diagnosis results of the SVM, DBN, LSTM and EATCN for these eleven fault patterns are established and exhibited in Figures 11 and 12. Specifically, Figure 11 provides the confusion matrices of these four fault diagnosis approaches. From Figure 11, the percentages in the dark orange blocks denote the percentages of accurately diagnosed fault samples, while the percentages in the shallow orange blocks stand for the percentages of mistakenly diagnosed fault samples. As displayed in Figure 11a–c, the percentages in the shallow orange blocks of the six rows for the SVM, DBN and LSTM are much greater than those in Figure 11d of the EATCN. This demonstrates that more fault data points pertaining to the fault AC are mistakenly identified by the SVM, DBN and LSTM. Moreover, in comparison with Figure 11d, many more shallow orange blocks appear in Figure 11a–c. This phenomenon means that the SVM, DBN and LSTM inaccurately discern many more fault data points of these eleven faults than that of the EATCN. To implement more graphical analysis and comparison, line charts of the index *FDR* values for the four approaches under the eleven fault patterns are exhibited in Figure 12. As revealed in Figure 12, the fault diagnosis rates

of the EATCN are significantly improved compared with those of the SVM, DBN and LSTM. To be specific, the EATCN's values of the index *FDR* for the eleven fault patterns are all above 92.00%, and the largest value of the index *FDR* even reaches 98.75%. As displayed in Figure 12, the great differences between the index *FDR*'s values of the four approaches prove the superiority of the EATCN for implementing the transmission line fault diagnosis.

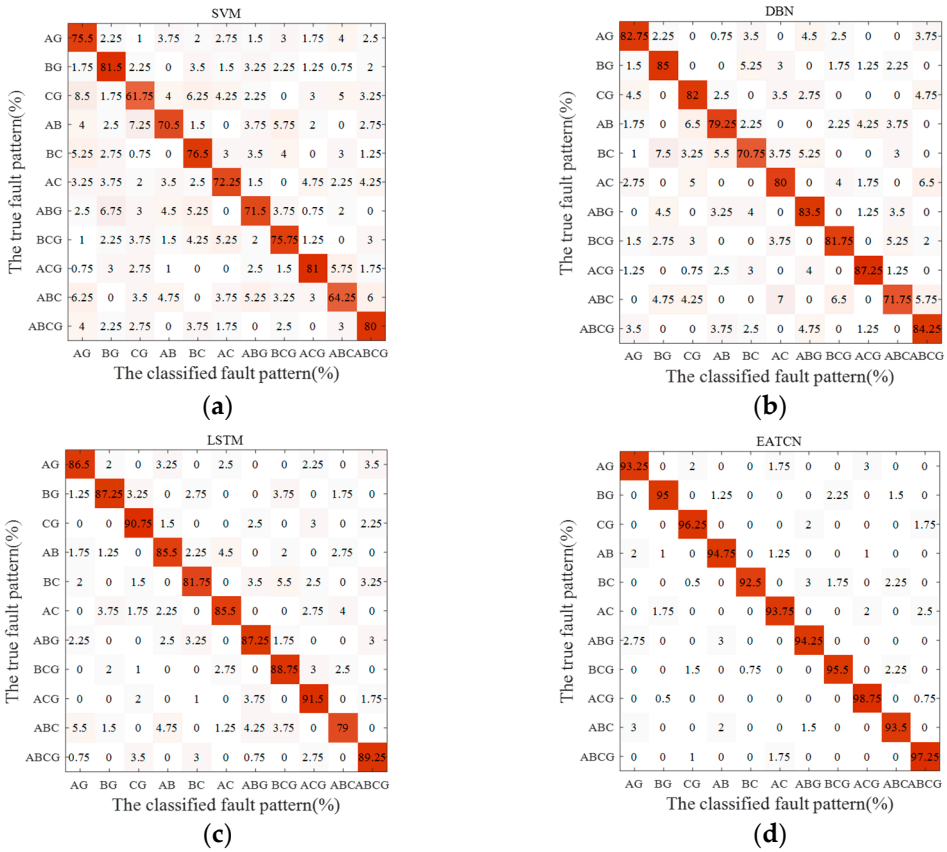


Figure 11. The confusion matrices of the SVM, DBN, LSTM and EATCN. (**a**) The confusion matrix of the SVM. (**b**) The confusion matrix of the DBN. (**c**) The confusion matrix of the LSTM. (**d**) The confusion matrix of the EATCN. The orange block is darker, the percentages of fault samples is more.

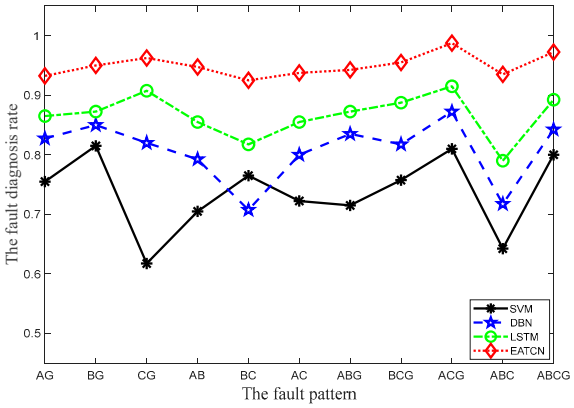


Figure 12. Line chart of the index *FDR*'s values for the four fault diagnosis approaches.

The fault diagnosis rates of the SVM, DBN, LSTM and EATCN for the eleven fault patterns are further quantized in Table 2. In addition, for the sake of fairness, the values of the index FDR_{average} for the four diagnosis models on the eleven fault patterns are also exhibited in Table 2. From Table 2, the values of the index FDR_{average} for the SVM, DBN, LSTM and EATCN are, respectively, computed as 73.68%, 80.75%, 86.64% and 94.98%.

Thus, the EATCN-based identification approach demonstrates the largest value of the index FDR_{average} for all eleven fault patterns among the four approaches, which testifies to the superiority of the EATCN's overall fault recognition effectiveness. Furthermore, in comparison with the SVM, DBN and LSTM, the suggested EATCN also exhibits more remarkable diagnosis performance to discern the particular fault of the eleven fault patterns. For example, the index *FDR*'s value of the fault pattern AB is 94.75% for the EATCN, in contrast to only 85.50% for the LSTM, 79.25% for the DBN and 70.50% for the SVM. Analogously, the value of the index *FDR* for the fault pattern ABCG is 97.25% for the EATCN, in comparison with only 89.25% for the LSTM, 84.25% for the DBN and even 80.00% for the SVM. It can be concluded that the presented EATCN approach is excellent at recognizing the short circuit fault patterns of the transmission line. This is because the global and local structure features extracted by the EATCN promote an improvement in the transmission line's fault identification task. To facilitate further visualized analysis, the values of the index *FDR* for the four algorithms under the eleven fault patterns are plotted in a histogram in Figure 13, which also proves the outstanding recognition performance of the EATCN over the SVM, DBN and LSTM for discerning all eleven short circuit faults.

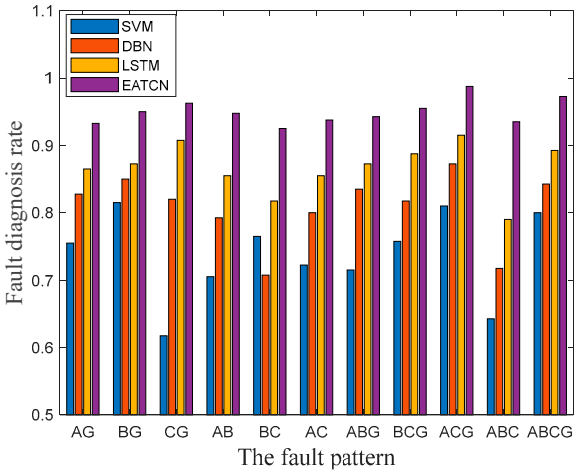


Figure 13. The identification results of the SVM, DBN, LSTM and EATCN for the eleven fault patterns.

Table 2. The values of the index *FDR* for the four fault diagnosis approaches.

Fault Pattern	SVM	DBN	LSTM	EATCN
AG	75.50%	82.75%	86.50%	93.25%
BG	81.50%	85.00%	87.25%	95.00%
CG	61.75%	82.00%	90.75%	96.25%
AB	70.50%	79.25%	85.50%	94.75%
BC	76.50%	70.75%	81.75%	92.50%
AC	72.25%	80.00%	85.50%	93.75%
ABG	71.50%	83.50%	87.25%	94.25%
BCG	75.75%	81.75%	88.75%	95.50%
ACG	81.00%	87.25%	91.50%	98.75%
ABC	64.25%	71.75%	79.00%	93.50%
ABCG	80.00%	84.25%	89.25%	97.25%
FDR_{average}	73.68%	80.75%	86.64%	94.98%

The values of the index precision *P* for the SVM, DBN, LSTM and EATCN are listed in Table 3. In addition, the values of the index P_{average} for the four diagnosis models are also exhibited in Table 3. From Table 3, the values of the index P_{average} for the SVM, DBN, LSTM and EATCN are, respectively, computed as 73.71%, 80.77%, 86.68% and 95.01%. Thus, the EATCN-based identification approach demonstrates the largest value of the index P_{average} for all eleven fault patterns, which testifies to the superiority of the EATCN's overall fault recognition effectiveness. In comparison with the SVM, DBN and LSTM, the suggested

EATCN displays a more remarkable diagnosis performance to discern the particular fault of the eleven fault patterns. For example, the index's precision value of the fault pattern ACG is 94.27% for the EATCN, in contrast to only 89.95% for the DBN, 84.92% for the LSTM and 82.03% for the SVM. Analogously, the value of the index precision for the fault pattern AB is 93.81% for the EATCN, in comparison with only 85.71% for the LSTM, 81.28% for the DBN and 75.40% for the SVM. It can be concluded that the presented EATCN is excellent at recognizing the short circuit fault patterns of the transmission line.

Table 3. The values of the index precision *P* for the four fault diagnosis approaches.

Fault Pattern	SVM	DBN	LSTM	EATCN
AG	66.96%	82.34%	86.50%	92.33%
BG	74.94%	79.63%	89.26%	96.69%
CG	68.04%	78.28%	87.47%	95.06%
AB	75.40%	81.28%	85.71%	93.81%
BC	72.51%	77.53%	86.97%	99.20%
AC	76.46%	79.21%	88.60%	95.18%
ABG	73.71%	79.71%	85.54%	93.55%
BCG	74.45%	82.78%	84.12%	95.98%
ACG	82.03%	89.95%	84.92%	94.27%
ABC	71.39%	79.06%	87.78%	93.97%
ABCG	74.94%	78.74%	86.65%	95.11%
P_{average}	73.71%	80.77%	86.68%	95.01%

6.4. Fault Diagnosis Effects of the Proposed EATCN under Different Noise Environments

To further verify the fault diagnosis effects of the proposed EATCN-based model under different noise environments, the Gaussian noise with zero mean and different variances is introduced to the monitored variables. The specific values of Gaussian noise's different variances are set to be 0.1, 0.01, 0.001 and 0.0001 through experience. In this way, the EATCN's fault diagnosis effects under different noise environments are tested and displayed in Tables 4 and 5.

Table 4. The values of the index *FDR* for the EATCN under different noise variances.

Fault Pattern	Variance (0.1)	Variance (0.01)	Variance (0.001)	Variance (0.0001)
AG	90.50%	93.25%	94.50%	95.25%
BG	91.25%	95.00%	97.25%	98.00%
CG	92.75%	96.25%	98.25%	98.75%
AB	90.75%	94.75%	96.00%	96.75%
BC	91.00%	92.50%	93.75%	94.50%
AC	90.25%	93.75%	95.50%	96.25%
ABG	91.50%	94.25%	96.25%	97.00%
BCG	93.50%	95.50%	97.50%	98.25%
ACG	98.00%	98.75%	100.00%	100.00%
ABC	89.50%	93.50%	95.00%	96.00%
ABCG	95.50%	97.25%	98.00%	98.50%
FDR_{average}	92.23%	94.98%	96.55%	97.20%

To be specific, Table 4 lists the EATCN's *FDR* and FDR_{average} values, while Table 5 exhibits the EATCN's indices *P* and P_{average} values for the eleven fault patterns, with the noise variance varying from 0.1 to 0.0001. When the value of noise variance is 0.1, which is the largest in our experiment, the EATCN achieves the worst fault diagnosis performance as the values of the FDR_{average} and P_{average} are both the smallest, i.e., 92.23% and 92.61%, respectively. However, the EATCN's values of FDR_{average} and P_{average} at the largest noise variance environment can be acceptable because they are both above 92.00%. With the decrement in the noise variance, the EATCN's fault diagnosis effect becomes better and better. But, when the noise variance decreases from 0.001 to 0.0001, the diagnosis effectiveness of the EATCN improves slightly because the FDR_{average} only varies

from 96.55% to 97.20% and the P_{average} only increases from 96.71% to 97.55%. Based on the above analysis, the developed EATCN shows outstanding accuracy and robustness under different noise environments.

Table 5. The values of the index precision P for the EATCN under different noise variances.

Fault Pattern	Variance (0.1)	Variance (0.01)	Variance (0.001)	Variance (0.0001)
AG	89.80%	92.33%	93.32%	94.52%
BG	93.29%	96.69%	97.58%	98.36%
CG	90.97%	95.06%	96.79%	97.53%
AB	92.51%	93.81%	95.56%	96.27%
BC	96.23%	99.20%	99.50%	100.00%
AC	93.07%	95.18%	96.74%	97.68%
ABG	91.83%	93.55%	95.98%	96.39%
BCG	94.42%	95.98%	97.70%	98.48%
ACG	91.46%	94.27%	97.26%	98.36%
ABC	92.08%	93.97%	96.53%	97.38%
ABCG	93.00%	95.11%	96.85%	98.13%
P_{average}	92.61%	95.01%	96.71%	97.55%

7. Conclusions

A novel EATCN fault recognition approach based on the CFP scheme and SCA mechanism is developed to identify the fault of the transmission line. To the best of our knowledge, this is the first study to incorporate the global and local structure feature extraction technique and an improved attention mechanism into the TCN model to diagnose the transmission line's fault. The other two contributions are expressed below. Firstly, a novel CFP-based dimension reduction technique is proposed to mine the global and local structure features of the transmission line data by integrating the local structure-preserving approach into the PCA model. Secondly, after the global and local structure features of the snapshot dataset are caught by using the established CFP technique, an EATCN-based diagnosis method is then suggested to classify these extracted features. Specifically, to cope with the gradient vanishing problem and improve the learning capability, the SCA network is first developed by incorporating a skip connection structure and two fully connected layers into the traditional attention mechanism. The attention-based TCN model is further constructed to dynamically pay different levels of attention to the imported global and local structure features by combining the developed SCA network with the conventional TCN's residual blocks. The detailed experiments certify the excellent fault diagnosis effect of the presented CFP- and SCA-based EATCN fault recognition strategy to diagnose the transmission line fault.

However, the adaptability and practical feasibility of the developed EATCN model are not assessed in our current work. One of our future works will be to assess the adaptability and practical feasibility of other new EATCN versions by verifying the fault diagnosis performance on the transmission line's experimental datasets collected from different sources or regions. In addition, in some practical applications, there is often a lack of labeled historical fault datasets, and it is difficult and expensive to manually label a large number of fault samples. Therefore, in order to extrapolate the suggested EATCN to other similar fault diagnosis tasks with inadequate training data, our future work will infuse the feature-based transfer learning methods [38,39] into the EATCN model to learn the domain invariant features by mapping the features of the source and target domains into a common feature space, allowing the data features of two domains to have distribution differences.

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

Funding: This research received no external funding.

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

Nomenclature

PCA	principal component analysis
LSP	local structure-preserving
CFP	comprehensive feature preserving
TCN	temporal convolutional network
SCA	skip connection attention
EATCN	enhanced feature extraction-based attention TCN
CNN	convolutional neural network
LSTM	long short-term memory
VRF	variable refrigerant flow
LG	line to ground
LL	double lines (line-to-line)
LLG	double lines to ground
LLL	triple lines
LLLG	triple lines to ground
PC	principal component
GRU	gate recurrent unit
ReLU	rectified linear unit
LSP	local structure-preserving
ATCN	attention temporal convolutional network
NF	no fault (normal operation)
AG	short fault of line a to ground
BG	short fault of line b to ground
CG	short fault of line c to ground
AB	short fault of line a to line b
BC	short fault of line b to line c
AC	short fault of line a to line c
ABG	short fault of line a and line b to ground
BCG	short fault of line b and line c to ground
ACG	short fault of line a and line c to ground
ABC	short fault of line a, line b and line c
ABCG	short fault of line a, line b and line c to ground
SVM	support vector machine
DBN	deep belief network
$\boldsymbol{X}$	original high-dimensional dataset
$\boldsymbol{x}(i)$	the i-th sample of the matrix $\boldsymbol{X}$
$\overline{\boldsymbol{x}}$	mean value of the samples
$\boldsymbol{x}(j)$	nearest neighbors of the $\boldsymbol{x}(i)$
$\boldsymbol{X}_{\text{sub}}(i)$	local neighborhood dataset subset of the $\boldsymbol{x}(i)$
n	the number of samples
m	the number of measured variables
$\boldsymbol{S}$	covariance of the datasets of the PCA
$\boldsymbol{D}$	diagonal matrix of the PCA
$\widetilde{\boldsymbol{P}}$	loading matrix of the PCA
$\boldsymbol{T}$	score matrix of the PCA
$\boldsymbol{t}_i$	the i-th score vector of the matrix $\boldsymbol{T}$
l	the number of retained leading vectors of the PCA
$\boldsymbol{E}$	residual matrix in PCA's the residual space
$\boldsymbol{F}(n)$	convolution computation of the input vector's n-th element
$\boldsymbol{x}_t$	input vector of the TCN
q	filter with the size of k
d	dilation coefficient

$\boldsymbol{p}$	loading vector of the CFP
$\boldsymbol{W}$	similarity matrix of the LSP
$\boldsymbol{W}_{ij}$	element of the similarity matrix $\boldsymbol{W}$
$w\{\boldsymbol{x}(i), \boldsymbol{x}(j)\}$	neighborhood relationship between the samples $\boldsymbol{x}(i)$ and $\boldsymbol{x}(j)$
$\boldsymbol{L}$	Laplacian matrix of the LSP
$J(\boldsymbol{p})_{PCA}$	objective function of the PCA
$J(\boldsymbol{p})_{LSP}$	objective function of the LSP
$J(\boldsymbol{p})_{CFP}$	objective function of the CFP
η	tradeoff parameter of the CFP
$\lambda_1, \lambda_2, \cdots, \lambda_d$	first d largest eigenvalues of the CFP
$\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_d$	eigenvectors of related to $\lambda_1, \lambda_2, \cdots, \lambda_d$ in the CFP
$\boldsymbol{P}$	loading vector of the CFP
$\boldsymbol{Y}$	score matrix of the CFP
$\tilde{\boldsymbol{X}}$	residual matrix of the CFP
$\boldsymbol{x}_F$	fault sample
$\boldsymbol{y}_F$	latent significant features
$\boldsymbol{X}_F$	snapshot dataset
$\boldsymbol{V}$	value vector of the SCA
$\boldsymbol{K}$	key vector of the SCA
$\boldsymbol{Q}$	query vector of the SCA
$\mathrm{F}_{\mathrm{score}}$	attention scores of the SCA
$FDR(i)$	fault diagnosis rate of the i-th fault pattern
FDR_{average}	average fault diagnosis rate

References

1. Liu, Y.; Lu, D.; Vasilev, S.; Wang, B.; Lu, D.; Terzija, V. Model-based transmission line fault location methods: A review. *Int. J. Electr. Power Energy Syst.* **2023**, *153*, 109321. [CrossRef]
2. Rajesh, P.; Kannan, R.; Vishnupriyan, J.; Rajani, B. Optimally detecting and classifying the transmission line fault in power system using hybrid technique. *ISA Trans.* **2022**, *130*, 253–264. [CrossRef]
3. Belagoune, S.; Bali, N.; Bakdi, A.; Baadji, B.; Atif, K. Deep learning through LSTM classification and regression for transmission line fault detection, diagnosis and location in large-scale multi-machine power systems. *Measurement* **2021**, *3*, 109330. [CrossRef]
4. França, I.A.; Vieira, C.W.; Ramos, D.C.; Sathler, L.H.; Carrano, E.G. A machine learning-based approach for comprehensive fault diagnosis in transmission lines. *Comput. Electr. Eng.* **2022**, *101*, 108107. [CrossRef]
5. Moradzadeh, A.; Teimourzadeh, H.; Mohammadi-Ivatloo, B.; Pourhossein, K. Hybrid CNN-LSTM approaches for identification of type and locations of transmission line faults. *Int. J. Electr. Power Energy Syst.* **2022**, *135*, 107563. [CrossRef]
6. Haq, E.U.; Jianjun, H.; Li, K.; Ahmad, F.; Banjerdpongchai, D.; Zhang, T. Improved performance of detection and classification of 3-phase transmission line faults based on discrete wavelet transform and double-channel extreme learning machine. *Electr. Eng.* **2021**, *103*, 953–963. [CrossRef]
7. Li, C.; Shen, C.; Zhang, H.; Sun, H.; Meng, S. A novel temporal convolutional network via enhancing feature extraction for the chiller fault diagnosis. *J. Build. Eng.* **2021**, *42*, 103014. [CrossRef]
8. Fahim, S.R.; Muyeen, S.M.; Mannan, M.A.; Sarker, S.K.; Das, S.K.; Al-Emadi, N. Uncertainty awareness in transmission line fault analysis: A deep learning based approach. *Appl. Soft Comput.* **2022**, *128*, 109437. [CrossRef]
9. Liu, L.; Liu, J.; Wang, H.; Tan, S.; Yu, M.; Xu, P. A multivariate monitoring method based on kernel principal component analysis and dual control chart. *J. Process Control.* **2023**, *127*, 102994. [CrossRef]
10. Zhu, A.; Zhao, Q.; Yang, T.; Zhou, L.; Zeng, B. Condition monitoring of wind turbine based on deep learning networks and kernel principal component analysis. *Comput. Electr. Eng.* **2023**, *105*, 108538. [CrossRef]
11. You, L.X.; Chen, J. A variable relevant multi-local PCA modeling scheme to monitor a nonlinear chemical process. *Chem. Eng. Sci.* **2021**, *246*, 116851. [CrossRef]
12. Lu, X.; Long, J.; Wen, J.; Fei, L.; Zhang, B.; Xu, Y. Locality preserving projection with symmetric graph embedding for unsupervised dimension reduction. *Pattern Recognit.* **2022**, *131*, 108844. [CrossRef]
13. Zhou, Y.; Xu, K.; He, F.; He, D. Nonlinear fault detection for batch processes via improved chordal kernel tensor locality preserving projections. *Control. Eng. Pract.* **2020**, *101*, 104514. [CrossRef]
14. He, Y.L.; Zhao, Y.; Hu, X.; Yan, X.N.; Zhu, Q.X.; Xu, Y. Fault diagnosis using novel AdaBoost based discriminant locality preserving projection with resamples. *Eng. Appl. Artif. Intell.* **2020**, *91*, 103631. [CrossRef]
15. Chen, B.J.; Li, T.M.; Ding, W.P. Detecting deepfake videos based on spatiotemporal attention and convolutional LSTM. *Inf. Sci.* **2022**, *601*, 58–70. [CrossRef]
16. Chen, Y.; Lv, Y.; Wang, X.; Li, L.; Wang, F.Y. Detecting traffic information from social media texts with deep learning approaches. *IEEE Trans. Intell. Transport. Syst.* **2018**, *20*, 3049–3058. [CrossRef]

17. Zhang, H.; Yang, W.; Yi, W.; Lim, J.B.; An, Z.; Li, C. Imbalanced data based fault diagnosis of the chiller via integrating a new resampling technique with an improved ensemble extreme learning machine. *J. Build. Eng.* **2023**, *70*, 106338. [CrossRef]
18. Guo, D.F.; Zhong, M.Y.; Ji, H.Q.; Liu, Y.; Yang, R. A hybrid feature model and deep learning based fault diagnosis for unmanned aerial vehicle sensors. *Neurocomputing* **2018**, *319*, 155–163. [CrossRef]
19. Lu, W.; Li, Y.; Cheng, Y.; Meng, D.; Liang, B.; Zhou, P. Early fault detection approach with deep architectures. *IEEE. Trans. Instrum. Meas.* **2018**, *67*, 1679–1689. [CrossRef]
20. Guo, Y.; Tan, Z.; Chen, H.; Li, G.; Wang, J.; Huang, R.; Liu, J.; Ahmad, T. Deep learning-based fault diagnosis of variable refrigerant flow air-conditioning system for building energy saving. *Appl. Energy* **2018**, *225*, 732–745. [CrossRef]
21. Zhou, Z.; Li, G.; Wang, J.; Chen, H.; Zhong, H.; Cao, Z. A comparison study of basic data-driven fault diagnosis methods for variable refrigerant flow system. *Energy Build.* **2020**, *224*, 110232. [CrossRef]
22. Greff, K.; Srivastava, R.K.; Koutník, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A search space odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *28*, 2222–2232. [CrossRef]
23. Guo, Q.; Zhang, X.H.; Li, J.; Li, G. Fault diagnosis of modular multilevel converter based on adaptive chirp mode decomposition and temporal convolutional network. *Eng. Appl. Artif. Intell.* **2022**, *107*, 104544. [CrossRef]
24. Dudukcu, H.V.; Taskiran, M.; Taskiran, Z.G.; Yildirim, T. Temporal convolutional networks with RNN approach for chaotic time series prediction. *Appl. Soft Comput.* **2023**, *133*, 109945. [CrossRef]
25. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, *452*, 48–62. [CrossRef]
26. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
27. Deng, D.; Jing, L.; Yu, J.; Sun, S. Sparse attention LSTM for sentiment lexicon construction. *IEEE/ACM Trans. Audio Speech Lang. Process* **2019**, *27*, 1777–1790. [CrossRef]
28. Nath, A.G.; Udmale, S.S.; Raghuwanshi, D.; Singh, S.K. Structural rotor fault diagnosis using attention-based sensor fusion and transformers. *IEEE Sens. J.* **2021**, *22*, 707–719. [CrossRef]
29. Fahim, S.R.; Sarker, Y.; Sarker, S.K.; Sheikh, M.R.; Das, S.K. Attention convolutional neural network with time series imaging based feature extraction for transmission line fault detection and classification. *Electr. Power Syst. Res.* **2020**, *187*, 106437. [CrossRef]
30. Fathabadi, H. Novel filter based ANN approach for short-circuit faults detection, classification and location in power transmission lines. *Int. J. Electr. Power Energy Syst.* **2016**, *74*, 374–383. [CrossRef]
31. Farshad, M.; Sadeh, J. Fault locating in high-voltage transmission lines based on harmonic components of one-end voltage using random forests. *Iran. J. Electr. Electron. Eng.* **2013**, *9*, 158–166.
32. Mitra, S.; Mukhopadhyay, R.; Chattopadhyay, P. PSO driven designing of robust and computation efficient 1D-CNN architecture for transmission line fault detection. *Expert Syst. Appl.* **2022**, *210*, 118178. [CrossRef]
33. Abdullah, A. Ultrafast Transmission Line Fault Detection Using a DWT-Based ANN. *IEEE Trans. Ind. Appl.* **2018**, *54*, 1182–1193. [CrossRef]
34. Howard, A.; Sandler, M.; Chu, G.; Chen, L.C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for mobilenetv3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1314–1324.
35. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
36. Fahim, S.R.; Sarker, S.K.; Muyeen, S.; Das, S.K.; Kamwa, I. A Deep Learning Based Intelligent Approach in Detection and Classification of Transmission Line Faults. *Int. J. Electr. Power Energy Syst.* **2021**, *133*, 107102. [CrossRef]
37. Kundur, P. *Power System Stability and Control*; McGraw-Hill: New York, NY, USA, 2022.
38. Kim, J.; Lee, J. Instance-based transfer learning method via modified domain-adversarial neural network with influence function: Applications to design metamodeling and fault diagnosis. *Appl. Soft Comput.* **2022**, *123*, 108934. [CrossRef]
39. Zhao, Z.B.; Zhang, Q.Y.; Yu, X.L.; Sun, C.; Wang, S.B.; Yan, R.Q.; Chen, X.F. Applications of unsupervised deep transfer learning to intelligent fault diagnosis: A survey and comparative study. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3525828. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.