



## DYNAMICALLY ENHANCED WEIGHTED NETWORK ACOUSTIC AND PHOTOELECTRIC GIS FAULT DIAGNOSIS BASED ON ATTENTION MECHANISM

Hui FU <sup>1</sup> , Yongling LU <sup>1</sup> , Zhen WANG <sup>1</sup> , Yujie LI <sup>1</sup> , Duohu GONG <sup>2</sup> , Weihao SUN <sup>3,\*</sup> 

<sup>1</sup> State Grid Jiangsu Electric Power Co., Ltd. Electric Power Science Research Institute, Nanjing 211100, China

<sup>2</sup> Xinjiang Electric Power Research Institute, State Grid Corporation of China, Urumqi 830000, China

<sup>3</sup> School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

\* Corresponding author, e-mail: [weihaosun250@163.com](mailto:weihaosun250@163.com)

### Abstract

In power systems, the normal functioning of gas-insulated switchgear (GIS) is essential for the security of the electrical grid. However, when a single signal is used for discharge detection and diagnosis, it will be interfered. Through joint analysis of different signals, fault diagnosis can be more accurately performed. Therefore, to address this problem, this paper proposes a dynamically enhanced weighted network model (AMB-DEWNM) based on the attention mechanism. The model first extracts fault features from the PRPD spectra of UHF, optical and ultrasonic signals through a multi-scale convolutional neural grid. Furthermore, a two-tier focus module is introduced to enhance fault characteristics that are insensitive to changes in operating conditions. Finally, a new dynamic enhanced weighted voting strategy (DEWVS) is designed. This strategy constructs a diagnostic performance index matrix by considering the diagnostic accuracy and misclassification rate of the base model to dynamically adjust the voting weight of each base model. distribution to obtain more reliable collaborative diagnostic results. Test outcomes demonstrate that the error detection precision of the AMB-DEWNM system is notably enhanced. Compared with other advanced network models, the diagnosis accuracy is as high as 95.28%. It has high stability and robustness, and provides fault detection and maintenance for GIS. strong support.

Keywords: power transformer; photoacoustic spectroscopy; domain adaptive intermediate class distribution alignment network; AdaSoftmax; fault diagnosis

## 1. INTRODUCTION

Power gas insulated switchgear (GIS) possesses the benefits of minimal upkeep expense and compact size, and is broadly applied in the sector of high-voltage power transmission. However, during the manufacturing, transportation, and assembly processes of GIS, internal insulation defects may occur, causing equipment failure [1, 2]. Partial discharge is an important basis for equipment status assessment and fault warning. Therefore, the detection and identification of partial discharge in GIS can promptly detect its internal insulation defects and ensure the safe operation of the equipment [3, 4].

When partial discharge occurs inside the equipment, electrical, magnetic, acoustic, light and other signals will be radiated outward. Based on the electromagnetic waves and acoustic vibrations caused by discharge, researchers have proposed ultra-high frequency (UHF) and acoustic emission

online detection methods [5, 6, 7]. However, there are complex and uncertain noises in the field, and both detection methods have certain challenges in collecting partial discharge signals [8]. Photometry is a new detection method for measuring and analyzing discharge signals. It has strong anti-interference and high detection confidence, and has broad development prospects and application value [9, 10, 11].

As optical fiber and photovoltaic transformation technologies advance, fluorescent optical fiber detectors, optical fiber acoustic sensors, and others have surfaced, enhancing the responsiveness of optical sensing techniques. Reference [9] introduced an interferometric partial discharge optical fiber acoustic detector, which improves GIS partial discharge detection sensitivity and maximum response amplitude compared with piezoelectric ceramic sensors. However, this article is an optical fiber sensing technique relying on sound release, that is fundamentally about measuring partial discharge

optical signals. Literature [12, 13] conducted experimental research and comparison on the characteristics of three fluorescent optical fiber sensors, which improved the reliability of motor stator insulation status monitoring. Literature [14, 15, 16] uses fluorescent optical fibers to obtain partial discharge light information in GIS, and proposes a multifractal spectral probability algorithm for calculating grayscale images of optical signals. Literature [17, 18, 19] used corona discharge in the air as the experimental object to study the wavelength alignment, connection extent, and positional relationship of luminescent light guides. In terms of partial discharge fault diagnosis, scholars have conducted extensive investigation and attained notable outcomes.

In order to solve the problem of incomplete single-source information, researchers use multi-source information composed of multiple sensor signals to obtain more complete equipment health status information, and build a multi-source information fusion model to obtain more stable and reliable diagnostic results. In the current fault diagnosis research based on multi-source information fusion, Deng et al. and Li et al. [20, 21] proposed a new intelligent diagnosis method based on multiple different sensor signal fusion and CNN; Bacha et al. [22] used vibration and acoustic signals to Features are extracted from the system and one-dimensional CNN is used to fuse features to achieve higher diagnostic accuracy; Liu et al. [23] proposed a one-dimensional CNN that fuses vibration, voltage and acoustic signals. fault diagnosis method, and used the Mahalanobis distance criterion to remove outliers in multi-source data; Wang et al., Rosenberg and Hirschberg, Hardoon et al. [24, 25, 26] applied wavelet packet decomposition to extract the time-frequency characteristics of current and vibration signals. Test outcomes indicate that the integrated characteristics yield Superior error detection precision.

In summary, this paper proposes a dynamically enhanced weighted network model (AMB-DEWNM) based on the attention mechanism. The model first extracts fault features from the PRPD spectra of UHF, optical and ultrasonic signals through a multi-scale convolutional neural grid. Additionally, a two-tier focus component is introduced to enhance fault characteristics that are insensitive to changes in operating conditions. Finally, a new dynamic enhanced weighted voting strategy (DEWVS) is designed. This strategy constructs a diagnostic performance index matrix by considering the diagnostic accuracy and misclassification rate of the base model to dynamically adjust the voting weight of each base model. distribution to obtain more reliable collaborative diagnostic results.

It can not only effectively handle partial discharge fault diagnosis tasks of GIS equipment, but also maintain high diagnostic accuracy and good generalization performance under complex and changeable actual working conditions. This

achievement provides strong technical support for GIS health management and fault warning, and is expected to further promote the intelligent operation and maintenance level of the power system.

## 2. AMB-DEWNM MODEL

### 2.1 Multi-scale feature extraction module

The architecture of the multi-level characteristic retrieval component is illustrated in Figure 1.

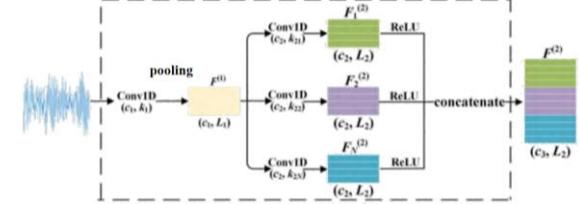


Fig. 1. Multi-level characteristic retrieval component

First, a wide convolution kernel is utilized to retrieve characteristics from the initial signal. The broad kernel acts as a low-pass filter. The convolved features can effectively reduce high-frequency noise and suppress data overfitting. The specific computation equation of the kernel operation is:

$$F^{(1)} = [f_1(k_1 * x + b); \dots; f_{c_1}(k_1 * x + b)] \quad (1)$$

Within the equation:  $k_1$  is the kernel matrix in the pathway;  $*$  is the kernel operator;  $x$  is the input data;  $b$  is the offset term of the kernel matrix;  $f()$  is the ReLU activation mechanism;  $c_1$  is the count of output pathways, and  $c_1$  each pathway generates the output characteristic  $F^{(1)}$ .

The output length in each channel after convolution is  $L_1$ , and the feature  $F^{(1)}$  size is  $(c_1, L_1)$ . Then, kernel layers with kernel matrices of varying dimensions are utilized to execute parallel processing on the output features of the previous layer, that is,  $N$  one-dimensional convolution layers convolve the features  $F^{(1)}$  at the same time. Each convolution layer will output a channel  $c_2$ , and the convolution layer will the convolution step size of the product layer is set to 1, and then activated by the ReLU activation function and then spliced along the channel axis. The total count of pathways in the ultimate output characteristic  $c_3 = N \times c_2$ . Utilizing kernel matrices of diverse dimensions to retrieve characteristics across distinct temporal scales greatly enriches fault feature information. Zero element padding is used during the convolution process to ensure that different convolution layers have the same output length  $L_2$ , so that the output features of different channels can be successfully spliced in the splicing layer. The calculation expression of the entire splicing process is:

$$F_i^{(2)} = f(k_{2i} * F^{(1)} + b) \quad (2)$$

$$F^{(2)} = \text{con}_i^N [F_i^{(2)}] \quad (3)$$

In the formula:  $k_{2i}$  is the kernel matrix applied in the  $i$ -th ( $i=1,2,\dots,N$ ) dimension;  $F^{(1)}$  is the input characteristic;  $b$  is the offset term of the kernel matrix;  $f()$  is the ReLU activation mechanism,  $\text{con}$  is

the splicing function;  $F_i^{(2)}$  is the  $i$ th features of  $i$  ( $i=1,2,\dots,N$ ) scales, the size of the feature  $F^{(2)}$  is  $(c_3, L_2)$ .

## 2.2 Dual-stage attention module

After each sensor signal passes through the multi-level characteristic retrieval component, the multi-level characteristics are weighted and adjusted from the pathway and spatial aspects by the two-tier focus component. The dual-stage attention module is shown in Figure 4. The dual-stage attention module is a key module to improve the model's adaptability to working conditions. It can enhance channel features that are more sensitive to bearing health status in the channel dimension and suppress channel features that are sensitive to changes in working conditions. It can enhance each channel feature in the spatial dimension. Features within the channel are more fault-discriminating and useless features are suppressed.

The detailed structure of the channel stage of the two-tier focus module is shown in Figure 2.

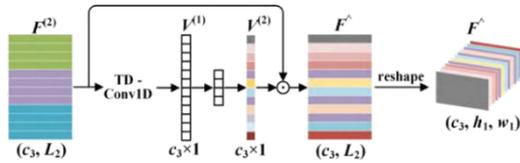


Fig. 2. Channel stage attention

The operation process of the channel stage: The difference from the classic SENet is that it does not use maximum pooling and average pooling to compress features, but uses a distributed one-dimensional convolution layer to cover all the features in each channel in order to achieve global convolution. The detailed description is shown in Figure 3.

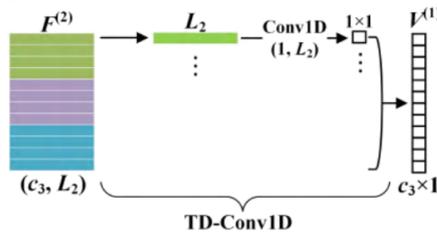


Fig. 3. Distributed 1D convolution

This process convolves the features within each channel into a single value in a vector, ultimately resulting in a single one-dimensional vector  $V^{(1)}$ . Then, a one-dimensional weight vector  $V^{(1)}$  is generated through the compression and excitation process, and the weight vector is multiplied element by element with the input features, that is, different weights are given to different channels to achieve the enhancement or suppression of different channel features. Its specific mathematical expression is:

$$V^{(2)} = \sigma(w_1 f(w_0(V^{(1)} + b_0) + b_1)) \quad (4)$$

$$F^\wedge = V^{(2)} \odot F^{(2)} \quad (5)$$

Among them,  $\sigma()$  is the sigmoid function;  $f()$  is the ReLU function;  $w_0, w_1$  is the coefficient of the densely linked level;  $b_0, b_1$  is the bias term;  $\odot$  is the element-wise multiplication;  $F^\wedge$  is the output feature of the channel stage attention.

The detailed structure of the spatial stage of the dual-stage attention module is shown in Figure 4.

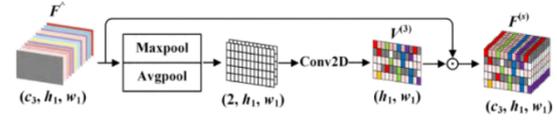


Fig. 4. Spatial stage attention

The operation process of the positional phase: initially execute max sampling and mean sampling procedures along the pathway axis, subsequently merge the two pathways, and apply a planar kernel operation with a kernel matrix dimension of 3 to calculate the spatial weight for weight distribution matrix  $V^{(3)}$ , this weight matrix encodes the position of the spatial feature to be enhanced or suppressed; finally, the feature  $F^\wedge$  and the generated spatial weight matrix  $V^{(3)}$  are multiplied element by element along the channel axis to achieve spatial feature weighting. Its specific mathematical expression is:

$$V^{(3)} = \sigma(\text{conv}^3([\text{Maxpool}(F^\wedge); \text{Avgpool}(F^\wedge)])) \quad (6)$$

$$F^{(s)} = V^{(3)} \odot F^\wedge \quad (7)$$

In the formula:  $\text{conv}^3$  represents the kernel operation featuring a kernel matrix dimension of 3; Maxpool is the max sampling operation; Avgpool is the mean sampling operation and  $F^{(s)}$  is the output feature of the spatial stage attention.

Finally, the weighted features are further extracted through planar kernel operation featuring a kernel matrix dimension of  $(k_{31}, k_{32})$  to mine deeper features as a whole; secondly, the batch normalization layer, pooling layer and Dropout layer are connected to improve the model. The purpose of generalization performance, reducing parameters and avoiding model overfitting; finally, the output features with size  $(c_4, h_2, w_2)$  will be flattened and then fully connected layers and softmax layers will be used to output the classification results.

## 2.3 Dynamically enhanced weighted voting strategy

In order to integrate all diagnostic results to provide more accurate and reliable collaborative diagnostic results. This paper uses a weighted voting method to achieve decision-making fusion. Since the distribution of weights is the key to weighted voting fusion, in order to obtain more accurate and reliable collaborative diagnosis results, a new dynamically enhanced weighted voting strategy is proposed, and its framework is shown in Figure 5.

Contrasted with alternative voting combination approaches, this strategy not only considers the diagnostic accuracy of each base model but also considers the misclassification rate of each base model when allocating weights. The details are as follows.

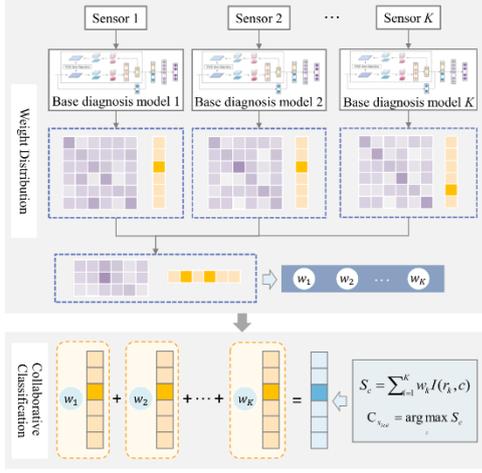


Fig. 5. Dynamic enhanced weighted voting strategy

Suppose data are collected from  $K$  sensors and there are  $C$  fault types to be identified.  $p_{ij}^{(k)}$  represent the probability that the  $k$ -th basic model determines the  $j$ -th fault type as the  $i$ -th fault type in the validation set. Then,  $p_{ij}^{(k)}$  will be calculated as

$$p_{ij}^{(k)} = \frac{F_{ij}^{(k)}}{N_j} \quad (8)$$

where  $N_j$  represents the overall count of instances of the  $j$ -th fault type in the validation set, and is the type of sample diagnosed as type  $i$  by the  $k$ -th basic model in the validation set  $j$  Number of samples. When  $i = j$ ,  $p_{ij}^{(k)}$  indicates the recognition accuracy from the  $k$ -th fundamental system for the  $j$ -th fault type. When  $i \neq j$ ,  $p_{ij}^{(k)}$  is the probability that from the  $k$ -th fundamental system misjudges the  $j$ -th fault type as the  $i$ -th fault type in the validation set.

For the test data  $x_{test}$ , it is assumed that the diagnostic results of the base model corresponding to  $K$  sensors are  $r_1, r_2, \dots, r_K$ , forming a set  $R_{test}$ . If the number of elements in the  $R$  test exceeds 1, it indicates that at least two of the underlying diagnostic models provide inconsistent diagnostic results. When assigning the weights of the  $k$ -th base model, the probability that the model is accurate and does not misclassify other fault types  $r_k$  used as an important indicator, and the weights are assigned as follows:

$$w_k = \frac{p_{r_k r_k}^{(k)} \prod_{r_j \in R_{test}, r_j \neq r_k} (1 - p_{r_j r_k}^{(k)})}{\sum_{i=1}^K p_{r_i r_i}^{(i)} \prod_{r_j \in R_{test}, r_j \neq r_i} (1 - p_{r_j r_i}^{(i)})} \quad (9)$$

The denominator plays the role of normalized weight.

For convenience, the matrix form of this process is given. First, construct a matrix  $A^{(k)}$  to record the diagnostic performance of the  $k$ -th basic model, expressed as a diagnostic performance index matrix, as follows:

$$A^{(k)} = (a_{ij})_{i,j-1}^C = \begin{bmatrix} a_{11} & \dots & a_{1C} \\ \vdots & \ddots & \vdots \\ a_{K1} & \dots & a_{KC} \end{bmatrix} \quad (10)$$

$$a_{ij} = \begin{cases} p_{ij}^{(k)}, & i = j \\ 1 - p_{ij}^{(k)}, & i \neq j \end{cases} \quad (11)$$

Then, calculate the diagnostic performance index matrix of the  $k$ -th basic model required for the test data as follows:

$$A_{test} = [r^{(1)T} A^{(1)}; \dots; r^{(C)T} A^{(C)}] \quad (12)$$

Where  $r^{(k)}$  is the one-hot vector of the prediction results of the  $k$ -th basic model. Therefore, the weight matrix is calculated as follows:

$$W = [w_1, w_2, \dots, w_K]^T = \frac{A_{test} r}{\|A_{test} r\|_1} \quad (13)$$

Here,  $r$  is the sign function result of the one-hot vector of all basic model pre-diagnosis results, that is:

$$r = \text{sgn} \left( \sum_{i=1}^K r^{(i)} \right) \quad (14)$$

where  $\|\cdot\|_1$  represents the norm  $l_1$  and  $\text{sgn}$  represents the symbolic function. After obtaining the weights of a series of basic diagnostic models, the score for each fault type is calculated by combining the diagnostic results of multiple sensors, as follows:

$$S_c = \sum_{i=1}^K w_k I(r_k, c), c = 1, 2, \dots, C \quad (15)$$

$$I(r_k, c) = \begin{cases} 1, & r_k = c \\ 0, & r_k \neq c \end{cases}$$

Finally, the fault type with the highest score is selected as collaborative fault diagnosis:

$$C_{xtest} = \arg \max S_c. \quad (16)$$

### 3. EXPERIMENTAL ANALYSIS

#### 3.1 AMB-DEWNM model verification analysis

To demonstrate the excellence of the AMB-DEWNM system, experiments were conducted. GIS PRPD patterns collected by UHF, ultrasonic, and optical sensors under four working conditions (C1-C4) were selected. In the experiment, GIS experienced various fault states, including free metal particle discharge (FP), suspended potential body discharge (FB), insulator surface discharge (SD), and metal tip discharge (MT). Each failure mode contains 409,600 PRPD patterns, which are organized into 100 samples; each sample contains 4,096 PRPD patterns. In the data set, each sample consists of 4096 PRPD patterns containing sufficient fault information, and the number of samples obtained under each fault is 1000. The dataset is split into learning set and validation set based on the proportion of 7:3.

Throughout learning, the iteration count is configured to 500. Considering that parameter settings have a significant impact on the model diagnosis results, the starting learning speed is  $\{0.1, 0.2, 0.3, 0.4, 0.5\}$ , and the learning rate attenuation value range is  $\{0.001, 0.002, 0.003, 0.004, 0.005\}$ , as shown in Figure 6, which depicts the impact of the

starting education speed and education speed decay factor on the system precision. Observations show that setting the initial education velocity to 0.1, and the education velocity reduction factor to 0.003, yields the optimal accuracy for the dataset. Therefore, this set of parameter configurations was selected.

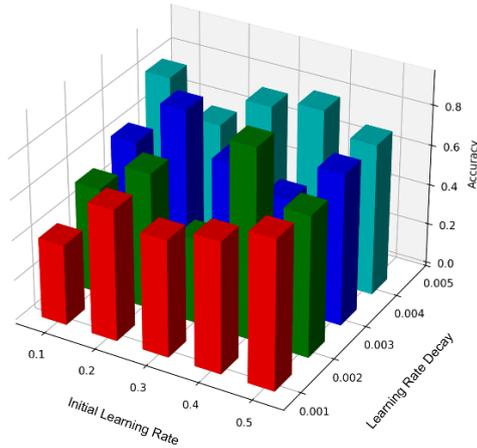


Fig. 6. 3D Bar Chart of initial learning rate and learning rate decay

To confirm the effectiveness of the system, compare it with the performance of the following neural network: DenseNet enhances the reusability of features and reduces the number of parameters through dense connections, but it consumes large memory and has high computational complexity. ResNet addresses the fading gradient issue in deep architectures via residual links and improves the training effect of the model. However, as the network deepens, the amount of parameters and computational complexity increase, making the model difficult to deploy in resource-constrained environments. Inception improves the expressive capacity of the system through multi-scale convolution and feature map splicing, but its structure is complex and difficult to implement, which is not conducive to rapid development and debugging. MobileNet significantly reduces the count of parameters and computations of the system via depthwise-separable kernel, and is particularly

suitable for deployment on mobile devices and embedded systems, but the performance may not be as good as more complex models in some complex tasks. ShuffleNet enhances the diversity of features through channel shuffle operations, improves the representation ability of the model, and maintains low computational complexity. However, it has certain requirements for hardware optimization, and its effect is limited when processing multi-source information fusion.

Figure 7 shows the training time, training accuracy and error corresponding to each neural network model under each task. Figure 8 shows the average training accuracy of each network under each task.

Experimental results show that DenseNet consumes a lot of memory and has high computational complexity, resulting in poor performance in real-time diagnosis tasks, especially in resource-constrained environments where it is difficult to deploy. In contrast, the AMB-DEWNM model uses multi-scale feature extraction and employs kernel matrices of various dimensions to retrieve characteristics across distinct temporal scales, which significantly improves the richness and robustness of features, while reducing computational complexity and is suitable for Real-time diagnostic tasks. ResNet As the network deepens, the number of parameters and computational complexity continue to increase. Deep networks are prone to overfitting problems, that influences the adaptability of the system.

The AMB-DEWNM model uses a dual-stage attention mechanism to weightedly modulate across the two aspects of pathway and spatial, enhances fault-sensitive features, significantly decreases the quantity of parameters and computational intricacy, and enhances the system's adaptability and robustness. Inception has a complex structure and high implementation difficulty, which is not conducive to rapid development and debugging. Multi-scale convolution and feature map splicing increase the complexity and computational burden of the model. AMB-DEWNM improves model

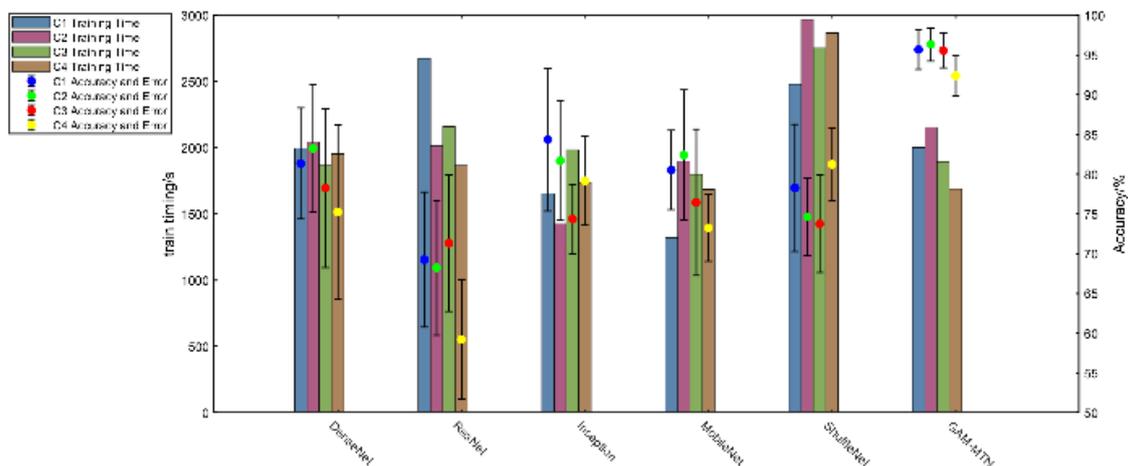


Fig. 7. Comparison of Diagnosis Accuracy and Training Time for Different Networks.

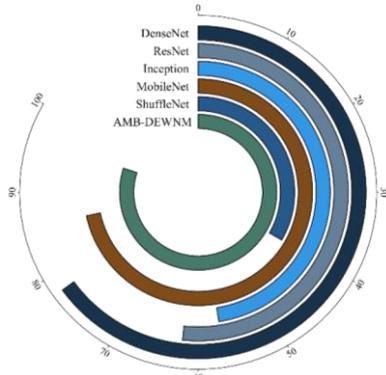


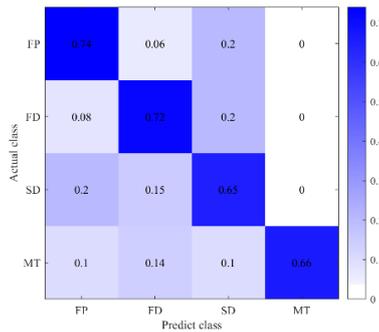
Fig. 8. Average Training Accuracy for Different Networks across Various Tasks

implementation and debugging efficiency by simplifying the structure while maintaining high performance. Although MobileNet greatly reduces the amount of parameters and calculations through depth-separable convolution, its performance is poor in certain complex tasks, especially multi-source information fusion and fault diagnosis tasks under complex working conditions. The AMB-DEWNM model passed, and the new dynamically enhanced weighted voting strategy dynamically allocated the

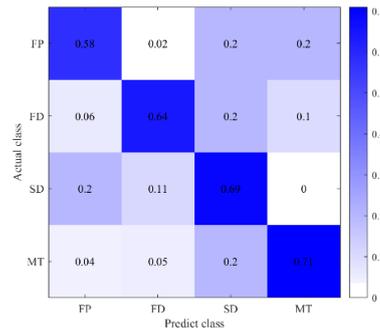
voting weights of each base model to further improve the accuracy and stability of diagnosis. ShuffleNet enhances the diversity of features and has limited effect when processing multi-source information fusion, especially its weak adaptive ability under different working conditions. AMB-DEWNM realizes the effective fusion of multi-source information, further improving the accuracy and stability of diagnosis without the need for complex optimization of hardware.

In order to study and understand the classification ability of the model in each category, the classification matrix for each system as depicted in Figure 9 was made based on experiments.

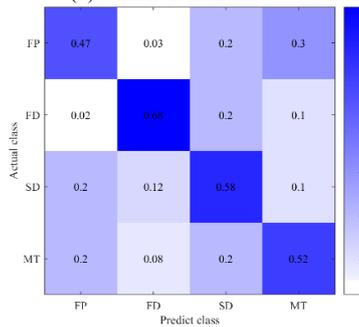
The confusion matrix shows the diagnostic accuracy of each model in different fault categories. It can be seen that AMB-DEWNM performs well in the diagnosis of most fault categories, especially in the diagnosis of SD and MT faults, achieving high accuracy. In contrast, other models perform poorly in diagnosing certain fault categories. ResNet has many errors in diagnosing FP faults, while Inception performs poorly in diagnosing FP faults. This suggests that these models may be inadequate in handling certain types of failures.



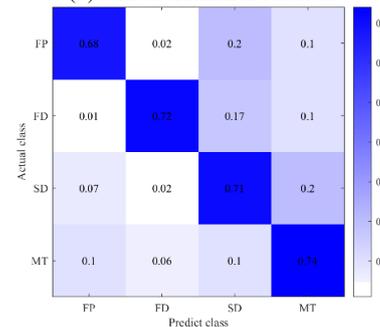
(a) DenseNet Confusion Matrices



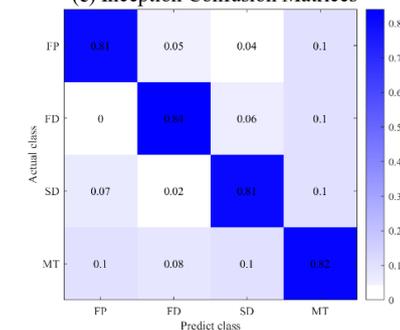
(b) ResNet Confusion Matrices



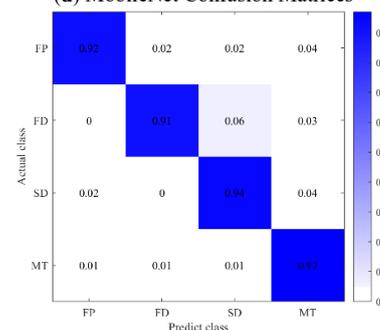
(c) Inception Confusion Matrices



(d) MobileNet Confusion Matrices



(e) ShuffleNet Confusion Matrices



(f) AMB-DEWNM Confusion Matrices

Fig. 9. Confusion Matrices for Different Networks

Through the confusion matrix, we can clearly see the advantages of AMB-DEWNM in diagnostic accuracy. It performs very well in the diagnosis of fault categories, which shows that AMB-DEWNM is an effective fault diagnosis tool.

#### 4. CONCLUSIONS

This paper proposes a dynamically enhanced weighted network model (AMB-DEWNM) based on the attention mechanism. The model first extracts fault features from the PRPD spectra of UHF, optical and ultrasonic signals through a multi-scale convolutional neural grid. Furthermore, a two-tier focus component is proposed to enhance fault characteristics that are insensitive to changes in operating conditions. Finally, a new dynamic enhanced weighted voting strategy (DEWVS) is designed. This strategy constructs a diagnostic performance index matrix by considering the diagnostic accuracy and misclassification rate of the base model to dynamically adjust the voting weight of each base model. distribution to obtain more reliable collaborative diagnostic results.

It can not only effectively handle partial discharge fault diagnosis tasks of GIS equipment, but also maintain high diagnostic accuracy and good generalization performance under complex and changeable actual working conditions. This achievement provides strong technical support for GIS health management and fault warning, and is expected to further promote the intelligent operation and maintenance level of the power system.

**Acknowledgment:** *This study was supported by Project Supported by the Science and Technology Project of State Grid Corporation of China (5500-202218132A-1-1-ZN).*

**Author contributions:** *mathematical model and the simulation techniques – H.F.; spelling and grammar checking as well as virtual validation – Y.L.; data interpretation and paper revision – Z.W.; data collection and data sorting – Y.L.; writing original draft preparation – D.G.; experimental validation – W.S.*

**Data availability statement:** *The datasets analyzed during the current study are available from the corresponding author on reasonable request.*

**Declaration of competing interest:** *The authors declare that they have no conflict of interest.*

#### REFERENCES

1. Wang YX, Yan J, Wang JH, Geng YS. A small-sample GIS partial discharge diagnosis method based on multi-level second-order attention twin networks, *Transac. China Electrotech. Soc.* 2023;38(8):2255-2264.
2. Hu ZY, Geng QY, Wei L, Chang YN, Li QM. The synergistic suppression effect and optimization design method of driving electrodes and particle traps in DC GIS/GIL, *Transac. China Electrotech. Soc.* 2023;38(12):3338-3349.
3. Du ZY, Hao ZY, Zhao PF, Wang H, Hao Q. Research on the propagation characteristics of ultrasonic guided waves in gas-insulated transmission lines suitable for sound source localization, *Transac. China Electrotech. Soc.* 2024;39(3):852-862.
4. Li Z, Wang H, Qian Y, Huang R, Cui QH. Noise-robust partial discharge pattern recognition based on accelerated robust features, *Transac. China Electrotech. Soc.* 2022;37(3):775-785.
5. Qu JL, Yu L, Yuan T, Tian YP, Gao F. An adaptive fault diagnosis algorithm for rolling bearings based on one-dimensional convolutional neural networks, *Chinese J. Sci. Instrument.* 2018;39(7):134-143. <https://doi.org/10.19650/j.cnki.cjsi.J1803286>
6. Ohenhen PE, Nwaobia NK, Nwasike CN, Gidiagba JO, Zni EC. Diagnostics and monitoring in electro-mechanical assemblies: assessing the latest tools and techniques for system health prediction. *Acta Electron. Malays.* 2024;8(1):11-20. <https://doi.org/10.26480/aem.01.2024.01.10>
7. Zhou X, Wu XT, Tian T, Bai J, Zhang ZY, Li JH. Condition assessment method of GIS disconnecter based on vibration-thermal-electrical multi-parameter signal characteristics, *Electric Power Engineering Technology.* 2024;43(1):220-228.
8. Liao JM, Guan XY, Lin JG, Liu J, Zhao JY. GIS vibration signal denoising and mechanical defect identification based on CycleGAN and CNN. *Electric Power Engineering Technology.* 2023;42(5):37-45. <https://doi.org/10.12158/j.2096-3203.2023.05.005>
9. Shi Z, Chen J, Zi Y, Chen Z. DecouplingNet: A stable knowledge distillation decoupling net for fault detection of rotating machines under varying speeds. *IEEE Trans. Neural Netw. Learn. Syst.* 2024;35(8):11276-11290. <https://doi.org/10.1109/TNNLS.2023.3258748>
10. Qian Q, Qin Y, Luo J, Wang Y, Wu F. Deep discriminative transfer learning network for cross-machine fault diagnosis. *Mech. Syst. Signal Process.* 2023;186:109884. <https://doi.org/10.1016/j.ymssp.2022.109884>
11. Qian Q, Zhou J, Qin Y. Relationship transfer domain generalization network for rotating machinery fault diagnosis under different working conditions. *IEEE Trans. Industr. Inform.* 2023;19(1):9898-9908. <https://doi.org/10.1109/TII.2022.3232842>
12. Feng Y, Chen J, He S, Pan T, Zhou Z. Globally localized multisource domain adaptation for cross-domain fault diagnosis with category shift. *IEEE Trans. Neural Netw. Learn. Syst.* 2023;34(6):3082-3096. <https://doi.org/10.1109/TNNLS.2021.3111732>
13. Wen L, Gao L, Li X. A new deep transfer learning based on sparse auto-encoder for fault diagnosis. *IEEE Trans. Syst., Man, Cybern., Syst.* 2019;49(1):136-144. <https://doi.org/10.1109/TSMC.2017.2754287>
14. Zhu J, Huang CG, Shen C, Shen Y. Cross-domain open-set machinery fault diagnosis based on adversarial network with multiple auxiliary classifiers. *IEEE Trans. Ind. Informat.* 2022;18(1):8077-8086. <https://doi.org/10.1109/TII.2021.3138558>
15. Jiao J, Zhao M, Lin J. Multi-weight domain adversarial network for partial-set transfer diagnosis. *IEEE Trans. Ind. Electron.* 2022;69(4):4275-4284. <https://doi.org/10.1109/TIE.2021.3076704>
16. Cao Z, Long M, Wang L, Jordan MI. Partial transfer learning with selective adversarial networks. in *Proc.*

- IEEE/CVF Conf. Comput. Vis. Pattern Recognit. 2018:2724-2732.
17. Li W, Chen Z, He G. A novel weighted adversarial transfer network for partial domain fault diagnosis of machinery. *IEEE Trans. Ind. Informat.* 2021;17(3): 1753-1762. <https://doi.org/10.1109/TII.2020.2994621>
  18. Xia Y, Shen C, Wang D, Shen Y, Huang W, Zhu Z. Moment matching-based intraclass multisource domain adaptation network for bearing fault diagnosis. *Mech. Syst. Signal Process.* 2022;168:108697. <https://doi.org/10.1016/j.ymssp.2021.108697>
  19. Chai Z, Zhao C, Huang B. Multisource-refined transfer network for industrial fault diagnosis under domain and category inconsistencies. *IEEE Trans. Cybern.* 2022;52(9):9784-9796. <https://doi.org/10.1109/TCYB.2021.3067786>
  20. Deng M, Deng A, Shi Y, Xu M. Correlation regularized conditional adversarial adaptation for multi-target-domain fault diagnosis. *IEEE Trans. Ind. Informat.* 2022;18(12):8692-8702. <https://doi.org/10.1109/TII.2022.3149906>
  21. Li J, Wang Y, Zi Y, Zhang Z. Whitening-net: A generalized network to diagnose the faults among different machines and conditions. *IEEE Trans. Neural Netw. Learn. Syst.* 2022;33(1):5845-5858. <https://doi.org/10.1109/TNNLS.2021.3071564>
  22. Bacha K, Souahlia S, Gossa M. Power transformer fault diagnosis based on dissolved gas analysis by support vector machine. *Electric Power Syst. Res.* 2012;83(1):73-79. <https://doi.org/10.1016/j.epsr.2011.09.012>
  23. Liu W, Wen Y, Yu Z, Li M, Raj B, Song L. SphereFace: Deep hypersphere embedding for face recognition. in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* 2017:212-220.
  24. Wang H. CosFace: Large margin cosine loss for deep face recognition. in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.* 2018:5265-5274.
  25. Rosenberg A, Hirschberg J. V-measure: A conditional entropybased external cluster evaluation measure. in *Proc. Joint Conf. Empirical Methods Natural Lang. Process. Comput. Natural Lang. Learn. (EMNLP-CoNLL)*. 2007:410-420. <https://doi.org/10.7916/D80V8N84>
  26. Haroon DR, Szedmak S, Shawe-Taylor J. Canonical correlation analysis: An overview with application to learning methods. *Neural Comput.* 2024;16(12): 2639-2664. <https://doi.org/10.1162/0899766042321814>



**Hui FU** received the Master's degree from the Shanghai Jiaotong University. She is currently working in State Grid Jiangsu Electric Power Co., Ltd. Electric Power Science Research Institute. Her current research interests including electric internet of things and artificial intelligence.  
e-mail: [wandering2000@163.com](mailto:wandering2000@163.com)



**Yongling LU** received the Master's degree from the Wuhan University. She is currently working in State Grid Jiangsu Electric Power Co., Ltd. Electric Power Science Research Institute. Her current research interests including electric internet of things and artificial intelligence.  
e-mail: [215105182955@163.com](mailto:215105182955@163.com)



**Zhen WANG** received the Master's degree from the South China University of Technology. He is currently working in State Grid Jiangsu Electric Power Co., Ltd. Electric Power Science Research Institute. His current research interests including electric internet of things and artificial intelligence.  
e-mail: [3951024011@qq.com](mailto:3951024011@qq.com)



**Yujie LI** received the Master's degree from the Xi'an Jiaotong University. He is currently working in State Grid Jiangsu Electric Power Co., Ltd. Electric Power Science Research Institute. His current research interests including partial discharge detection, switchgear status evaluation and fault analysis.  
e-mail: [41275911032@qq.com](mailto:41275911032@qq.com)



**Duohu GONG** received the Bachelor's degree from the Wuhan University. He is currently working in State Grid Xinjiang Electric Power Co., Ltd. Electric Power Science Research Institute. His current research interests including diagnostic analysis of switchgear.  
e-mail: [gongduohu@126.com](mailto:gongduohu@126.com)



**Weihao SUN** received the Bachelor's degree from the Harbin Institute of Technology. He is currently pursuing the Ph.D. degree in electrical engineering with Shanghai Jiao Tong University. His current research interests including online monitoring and fault diagnosis for power equipment insulation.  
e-mail: [weihaosun250@163.com](mailto:weihaosun250@163.com)