A Data Augmentation and Deep Residual Network–Based Method for Water Detection in Cable Elbow Terminations

Zuliang Yin[1]*, Yuxiang Zuo[2], Jianmin Lin[1], Zongguo Tao[1], Gang Zhou[1], Hua Chen[1], Tengfei Li[1]

[1]Xuancheng Power Supply Company, State Grid Anhui Electric Power Co., Ltd., Anhui, China
[2]Electric Power Research Institute, State Grid Anhui Electric Power Co., Ltd., Anhui, China

Corresponding author: Zuliang Yin (e-mail: yinzuliang0614@163.com).

ABSTRACT This paper uses an acoustic fingerprint recognition technique to solve the problems posed by lack of samples and degraded recognition performance in low signal-to-noise ratio (SNR) situations when detecting water ingress in cable elbow termination. This technique incorporates data augmentation and deep residual networks. Two unique data augmentation techniques, Speed Perturbation (SP) and Spectrogram Masking (SM), are proposed to extend sample diversity in the time and time-frequency domains. The classification backbone here is ResNet-18, which uses skip connections in order to address the issue of vanishing gradients, enabling the system to perform the recognition task from Mel spectrograms to states of water ingress in an end-to-end manner. Experimental evidence illustrates that ResNet-18 minimizes the total error rate from 9.68% (CNN baseline) to 3.46%. The SP+SM strategy showed the best results in the 0–5 dB high-noise region. In the 5–15 dB moderate-noise regions, SM alone showed the best results, while SP alone is better suited to clean, high-SNR scenarios. These findings provide a rationale for dynamically implementing augmentation strategies in practical settings. The non-contact and easy-to-deploy nature of this method presents a promising approach for the intelligent detection of water ingress hazards in power distribution equipment.

INDEX TERMS Cable elbow termination water-ingress; acoustic fingerprint recognition; Data Augmentation; Deep Residual Networks

I. INTRODUCTION

In the context of rapid socioeconomic development and continued urbanization, the construction of urban distribution networks is expanding. Due to their high supply reliability, compact construction, and good environmental adaptability, power cables have become the dominant transmission medium of the urban distribution networks at the medium and low voltages, and they are now widely used in 10 kV urban distribution systems [1], [2].

Cable elbow terminations are important accessories for cable connections in ring main units. They provide essential support for reliable interconnection and firm power transmission within the distribution system. Their operating status directly affects the entire power system reliability [3]. With prolonged working hours, cable elbow termination devices may suffer from a number of factors resulting in loss of sealing integrity, which allows water or moisture to penetrate the device. Some of these factors are improper installation, seal failure and aging, moisture ingress that occurs during high outdoor humidity, and external mechanical impacts [4], [5]. These internal defects are difficult to identify using normal inspection methods because moisture builds up very slowly and there are no obvious symptoms until significant damage occurs. Once moisture collects inside, it further deteriorates the insulation system, accelerates insulation aging, and increases the risk of partial discharges. These discharges can lead to total insulation failure and cause cascading outages. This presents significant challenges to the safe operation of urban distribution networks [6]. Therefore, developing techniques to detect moisture in cable elbow terminations is essential for condition monitoring and predictive maintenance of distribution systems [7].

Cable accessories may currently be diagnosed using a series of methodologies including: manual inspections, insulation resistance evaluations, withstand voltage assessments, partial discharge evaluations, infrared thermography, and ultrasonic inspections etc. [8]–[12]. While all these methods have been used frequently for fault diagnosis and condition assessments, they each have specific shortcomings in the diagnosis of water accumulation in cable elbow terminations. First, it is important to note that elbow terminations, which are critical components of cable accessories, are often mounted in closed interior cabinets or at the interface of the equipment meaning that internal water accumulation cannot

be identified using visual inspection. Second, some methods of electrical testing require the unit under test to be de-energized or require special test equipment, which, in turn, means that testing may become complex and is therefore unsuitable for expedient non-intrusive field diagnostics. Finally, during the initial stages of moisture intrusion, the signals associated with the phenomenon are often weak and easily masked by extraneous noise, variation in the installation and in acquisition conditions. Consequently, traditional approaches relying on hand-crafted features and fixed thresholds demonstrate poor adaptability and limited generalization.

The acoustic-based sensing technique allows for contactless implementation, is inexpensive, and can be deployed in a variety of field scenarios while providing a viable new approach for identifying water build up in cable elbow termination [13], [14]. Moisture or water build-up inside an elbow termination changes the dielectric properties, structural responses, and material behavior. These changes lead to alterations in the propagation of acoustics and the changes can be seen in the acquired acoustic data. Time-frequency representation, in contrast to the reliance on the time-domain amplitude, or attributes of frequency-domain peak characteristics, shows a more extensive explanation on the energy distribution and transient response behavior of different moisture states. Therefore, transforming acoustic data into time-frequency representation and combining that with a form of intelligent recognition approaches to automatic feature extraction, can improve the accuracy and certainty of identifying water build-up.

Recently, deep learning approaches have been used widely in voice recognition, mechanical fault diagnosis, and acoustic classification [15], [16]. Convolutional neural networks (CNNs) can learn hierarchical representations automatically from raw data, thereby minimizing the need for manual elaboration of attributes, common in previous methods [17], [18]. Deep residual networks (ResNet) specifically have introduced skip connections that alleviate the vanishing gradient and performance drops associated with deeper networks, allowing for greater high-level feature extraction.

The primary challenge of using raw acoustic signals to determine water accumulation in cable elbow terminations is the broad range of amplitude variations, unsteady energy distributions, inadequate sample sizes, and inconsistencies in the recording conditions. Because of the unreliability of raw signals, neural networks can have unsteady training or generalize poorly. Because of this, developing a sophisticated data processing system that incorporates a unique method for data preprocessing, augmentation, and feature extraction aimed at the detection of water accumulation via acoustic signatures is important to obtain reliable results.

To address the aforementioned challenges, this paper presents a novel method for detecting water accumulation in cable elbow terminations by integrating data augmentation with deep residual networks. The primary contributions of this work are threefold:

(1) Dual-Domain Data Augmentation Strategy: A combined data augmentation framework is proposed, incorporating Speed Perturbation (SP) in the time domain and Spectrogram Masking (SM) in the time-frequency domain. This dual-domain approach significantly enhances sample diversity and improves model robustness against both temporal variations and spectral information loss.
(2) ResNet-Based Acoustic Recognition for Water Detection: To the best of our knowledge, this is the first study to apply a deep residual network (ResNet-18) to the task of water ingress detection in cable elbow terminations using acoustic fingerprint recognition. The residual architecture enables effective end-to-end learning from Mel spectrograms to water ingress states, achieving substantially improved recognition accuracy and convergence stability compared to conventional CNN models.
(3) Noise-Adaptive Augmentation Strategy Selection: Through systematic experiments under varying signal-to-noise ratio (SNR) conditions, this study reveals the optimal augmentation strategy selection mechanism: SP+SM for low-SNR environments (0–5 dB), SM alone for moderate-SNR conditions (5–15 dB), and SP alone for high-SNR scenarios (> 15 dB). This finding provides practical guidance for dynamically configuring augmentation strategies in field deployments.

The remainder of this paper is organized as follows: Section II describes the overall framework for acoustic-based water accumulation detection, including the impact-based detection mechanism and signal processing pipeline. Section III presents the proposed data augmentation strategies in detail. Section IV elaborates on the deep residual network architecture. Section V reports comprehensive experimental results and comparative analyses. Finally, Section VI concludes the paper with a summary of findings and directions for future research.

II. ALGORITHM OVERVIEW

The cable elbow connector water ingress acoustic fingerprint recognition algorithm uses modular construction and can be broken down into three essential functional modules: signal processing, deep learning-based feature extraction, and state recognition. Figure 1 illustrates the algorithm's overall workflow.

Signal preprocessing module is shown in Fig. 2. It serves as the algorithm's front-end processing unit, where the acoustic signals captured from cable elbow terminals are subjected to speed perturbation augmentation in order to increase data variety and model robustness. This method is classified as time-domain data augmentation and involves changing the playback speed of an acoustic signal through time stretching while the pitch remains the same. This approach helps replicate the signal variations that may arise from differing sampling conditions and thus expands the training dataset effectively.

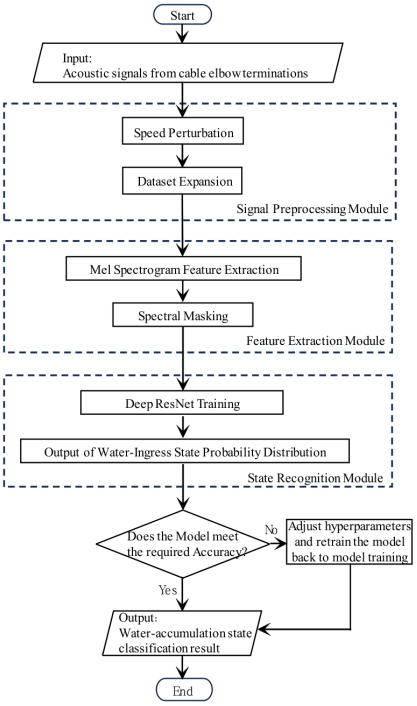

FIGURE 1. Overall Framework of Acoustic Fingerprint Recognition.

The feature extraction module is responsible for the time-frequency representation and modeling of the learned preprocessed acoustic signals. To do this, one-dimensional time-domain signals are converted into Mel-spectrograms, which capture the energy distribution aligned with human auditory perception. Additionally, the proposed model's generalization ability and robustness are further enhanced by the introduction of a Spectrogram Masking augmentation strategy, which randomly masks local regions in the spectrogram along both the time and frequency axes. This technique simulates partial signal loss and environmental interference that can occur during real-world data collection.

The module for state recognition is the main function for feature learning and decision-making. The classifier backbone is a deep residual network (ResNet) that performs end-to-end learning and classification of the augmented Mel-spectrogram features. Due to the use of residual connections, ResNet can circumvent the vanishing gradient problem which plagues many deep neural networks, and thus, is able to obtain better high-level discriminative semantic features. The proposed technique is built upon this framework, and thus, can accurately detect and classify the different water-ingress states of the elbow terminal cables.

III. DATA AUGMENTATION STRATEGY

A. SPEED PERTURBATION

Speed perturbation is a time domain data augmentation method which causes the input signal to vary temporally instead of directly resampling. It does this by changing the playback speed of the input signal to create the appearance of modified sampling rates.

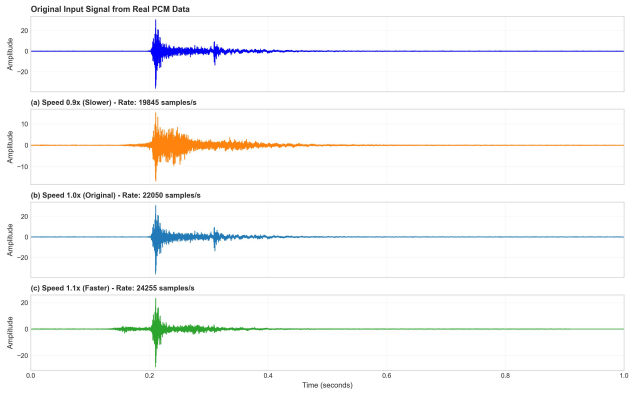

FIGURE 2. Example Samples with Speed Perturbation.

More specifically, speed perturbation on an original signal $x[n]$ of length N is applied using a scaling factor of $r \in \{0.9, 1.0, 1.1\}$ via the `time_stretch` method in the librosa library. This method conducts phase reconstruction in the output signal's frequency domain to keep the pitch of the signal unchanged. Thus, the length of the resulting augmented signal is $\lceil N/r \rceil$, which is then adjusted to the original length by either truncation or zero-padding.

Speed perturbation has a few advantages, as it can first enlarge a training set by increasing the total number of samples from N to $3N$. Second, it helps improve the robustness of the model against time-related variations by mimicking different sampling rates that could occur during real-world data collection. Additionally, it maintains the essential structural features of the acoustic signal, avoiding the potential information loss from standard resampling techniques.

Speed perturbation is applied to the raw PCM signal before it is converted into a Mel-spectrogram in the proposed framework. Time-domain augmentation allows the following spectrogram masking to be applied to the modified Mel-spectrogram. This way, it enables additional augmentation in the time and frequency domains. Figure 2 shows example waveforms produced by the speed perturbation method.

B. SPECTROGRAM MASKING AUGMENTATION

Spectrogram masking augmentation enhances the classifier's robustness to missing observations by randomly masking portions of the mel-spectrogram in the time-frequency domain. Let the input mel-spectrogram be defined as:

$$\mathbf{M} = [m_{f,t}] \in \mathbb{R}^{F \times T} \tag{1}$$

where F denotes the frequency dimension length (number of mel-frequency bins), T denotes the temporal dimension length, and $m_{f,t}$ represents the magnitude at the f-th frequency bin at the t-th time frame.

This work employs three masking strategies to augment the mel-spectrogram: frequency masking, time masking, and their combination.

Frequency Masking randomly selects a contiguous interval along the frequency dimension and sets it to zero. Given a frequency masking ratio α_f, the frequency masking width is defined as:

$$L_f = \lceil \alpha_f F \rceil \tag{2}$$

where $\lceil \cdot \rceil$ denotes the ceiling function. In this work, we set $\alpha_f = 0.2$.

The masking start position is uniformly randomly sampled from the valid range:

$$f_0 \sim U(0, F - L_f) \tag{3}$$

where $U(a, b)$ denotes the uniform distribution over $[a, b]$.

The frequency masking operation sets the selected frequency interval to silence (-80 dB):

$$M^{(f)}(f,t) = \begin{cases} -80, & f_0 \le f \le f_0 + L_f, \\ M(f,t), & \text{otherwise}, \end{cases} \quad \begin{matrix} f=1,2,\ldots,F, \\ t=1,2,\ldots,T \end{matrix} \tag{4}$$

where all spectral components within the masked frequency interval are set to zero across all time frames.

To enable unified representation, we introduce the frequency masking matrix, defined as:

$$A_f(f,t) = \begin{cases} -80, & f_0 \le f \le f_0 + L_f, \\ 1, & \text{otherwise}. \end{cases} \tag{5}$$

The frequency masking operation can then be expressed as:

$$\mathbf{M}^{(f)} = \mathbf{M} \odot \mathbf{A}_f \tag{6}$$

where $\odot$ denotes the Hadamard element-wise product.

Time Masking randomly selects a contiguous interval along the temporal dimension and sets it to zero. Given a temporal masking ratio α_t, the time masking width is defined as:

$$L_t = \lceil \alpha_t T \rceil \tag{7}$$

In this work, we set $\alpha_f = 0.1$.

The starting position of time masking is uniformly randomly sampled from the feasible interval:

$$t_0 \sim U(0, T - L_t) \tag{8}$$

The mel-spectrogram after time masking is then expressed as:

$$M^{(t)}(f,t) = \begin{cases} -80, & t_0 \le t \le t_0 + L_t, \\ M(f,t), & \text{otherwise}, \end{cases} \quad \begin{matrix} f=1,2,\ldots,F, \\ t=1,2,\ldots,T \end{matrix} \tag{9}$$

where all frequency components within the masked temporal interval are zeroed.

Similarly, we introduce the time masking matrix, defined as:

$$A_t(f,t) = \begin{cases} -80, & t_0 \le t \le t_0 + L_t, \\ 1, & \text{otherwise}. \end{cases} \tag{10}$$

The time masking operation is then expressed as:

$$\mathbf{M}^{(t)} = \mathbf{M} \odot \mathbf{A}_t \tag{11}$$

To simultaneously enhance the model's robustness to frequency-localized and temporally-localized information loss, we employ a combined masking approach. Frequency masking is first applied to the original mel-spectrogram, followed by time masking:

$$\mathbf{M}' = \mathbf{M} \odot \mathbf{A}_f \tag{12}$$

$$\mathbf{M}'' = \mathbf{M} \odot \mathbf{A}_t \tag{13}$$

where $\mathbf{A}_f$ and $\mathbf{A}_t$ denote the frequency and time masking matrices, respectively.

Since the two masking operations sample independently, the final augmented spectrogram can be expressed in unified form as:

$$\mathbf{M}^* = \mathbf{M} \odot \mathbf{A}_f \odot \mathbf{A}_t \tag{14}$$

Expanding this into piecewise form:

$$\mathbf{M}''(f,t) = \begin{cases} -80, & \begin{matrix} (f_0 \le f \le f_0 + L_f) \\ \text{or } (t_0 \le t \le t_0 + L_t), \end{matrix} \\ M(f,t), & \text{otherwise}, \end{cases} \quad \begin{matrix} f = 1, \ldots, F, \\ t = 1, \ldots, T \end{matrix} \tag{15}$$

As shown in Eq. (15), combined masking simultaneously introduces both frequency-localized and temporally-localized information loss in the time-frequency plane, thus constructing a more diverse set of augmented samples.

A key design principle of our approach is the randomness of masking positions. The masking positions $(f_{\text{start}}, t_{\text{start}})$ for each training sample are independently randomly sampled. This design offers three key advantages:

(1) Sufficient Data Augmentation. During training across multiple epochs, the model encounters a large variety of different masking combinations, enabling it to learn discriminative features under diverse constraint conditions.

(2) Prevention of Catastrophic Forgetting. Random masking prevents the model from memorizing the relationship between specific samples and specific masking patterns, thereby avoiding overfitting to any particular loss pattern.

(3) Implicit Regularization. This random constraint acts as a noise-based regularization mechanism, further constraining model complexity beyond standard dropout.

The three masking approaches are illustrated in Fig. 3.

The energy effects of frequency and time masking are illustrated in Figures 4 and 5.

Looking at Figure 4, the per-frequency energy comparison makes it clear that frequency masking can effectively cut down the energy within the masked frequency bands, while Figure 5 gives the per-time energy analysis and shows that time masking leads to quite noticeable energy attenuation inside the masked interval. Taken together, these two visualizations confirm that the masking operations do work as intended.

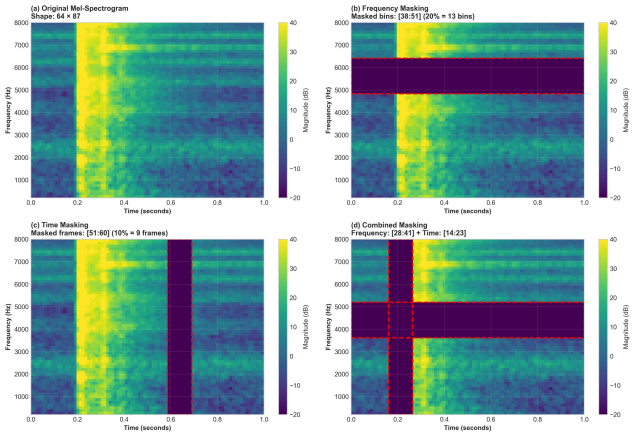

FIGURE 3. Examples of Three Masking Methods.

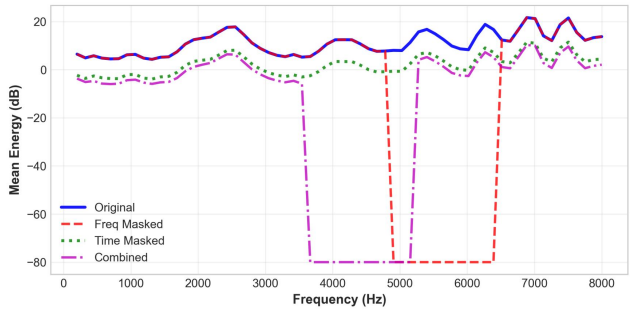

FIGURE 4. Effect of Frequency Masking on Energy Distribution.

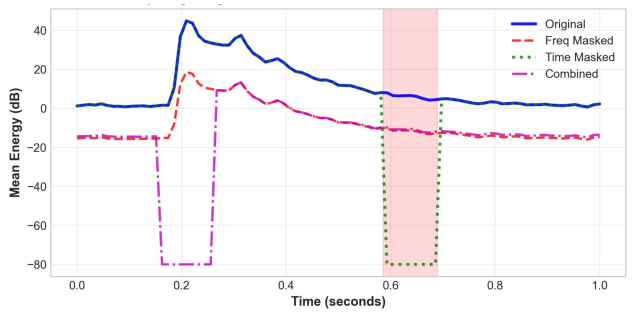

FIGURE 5. Effect of Frequency Masking on Energy Distribution.

IV. DEEP RESIDUAL NETWORK MODEL

A. FUNDAMENTAL PRINCIPLE OF RESIDUAL LEARNING

By making use of skip connections, the deep residual network (ResNet) gets around the vanishing gradient problem that often troubles conventional deep convolutional neural networks (CNNs). Its core idea is to learn a residual mapping, rather than trying to directly approximate the target mapping. Rather than learning the target function $H(x)$ directly, the network learns the residual function $F(x) = H(x) - x$, such that the final output can be expressed as

$$H(x) = F(x) + x \tag{16}$$

Residual learning brings a number of benefits: for one thing, skip connections allow gradients to flow directly through the network during backpropagation, which helps ease the gradient vanishing problem; for another, the residual architecture enables networks to be built much deeper without running into the degradation phenomenon, leading to stronger feature representation capability; in addition, low-level features learned by shallow layers can be passed straight to deeper layers, cutting down on redundant learning of fundamental features and boosting feature reuse.

Formally, a standard residual block can be expressed as

$$y = F(x; \{W_l\}) + W_s x \tag{17}$$

where $F(x; \{W_l\})$ denotes the nonlinear residual mapping learned by a series of convolutional layers parameterized by $\{W_l\}$, and W_s represents an optional linear projection used to match the dimensionality of the input with that of the residual branch when necessary. This residual formulation enables efficient optimization of deep neural networks while maintaining high feature extraction capability.

B. RESNET ARCHITECTURE DESIGN

Considering the requirements of the liquid level detection task, the selected primary feature learning model is the deep residual network ResNet-18. ResNet-18 consists of 18 convolutional layers, including one convolutional layer, four residual stages, and a classification layer, totaling approximately 11.2 million parameters. This architecture has distinct advantages over other, shallower CNNs. Firstly, the depth of the network enables it to learn feature representations hierarchically, transitioning from primitive time-frequency features to more refined and abstract semantic representations. Additionally, due to the network's depth, skip connections provide multiple pathways for the gradient to flow, which mitigates the vanishing gradient problem and ensures that every layer receives gradient updates. Finally, deep networks progressively downsample feature maps, resulting in an extensive receptive field that allows the network to identify long-range relationships within the liquid impact acoustic signals.

The first layer modifies the 6-channel Mel-spectrogram input into 64 feature channels and performs initial spatial downsampling using a 7×7 convolutional kernel with a stride of 2. Next, a batch normalization layer and ReLU activation are followed by a 3×3 max pooling layer (with stride $= 2$). These two consecutive operations with a stride of 2 reduce the spatial dimensions of the feature map from $(64, 281)$ to $(64, 71)$ while preserving 64 output channels. This design is intended to facilitate quick acquisition of global acoustic patterns to reduce depth of computation and provide a strong foundation for subsequent, more extensive feature extraction.

Residual Stage 1 (Layer 1) contains two standard residual blocks, each of which has two 3×3 convolution blocks, followed by batch norm and ReLU activation modules. In this stage, the input and output channel dimensions remain

64, and the stride is 1 for all scenarios so the spatial dimensions do not change. This allows the full time-frequency resolution to be preserved, which allows the network to learn the original spectrogram in fine detail. This level of detail is very important for liquid level detection.

Residual Stage 2 (Layer 2) also contains two residual blocks, but this time also increases both the number of channels and the size of the spatial dimension. The first block has a stride of 2, which increases the number of channels from 64 to 128 and decreases the size of the feature map from $(64, 71)$ to $(128, 36)$. The following block maintains the same dimensions. This downsampling increases the size of the receptive field at the expense of spatial resolution, which means more of the acoustic context can be captured in the deeper layers.

Residual Stage 3 (Layer 3) increases the capacity for feature representation even further. Again, there are two residual blocks; the first of which performs $2\times$ downsampling and increases the channel dimension from 128 to 256, while also reducing the feature map size from $(128, 36)$ to $(256, 18)$. The second block keeps these settings. The receptive field is now approximately $16\times$ larger than the input, allowing the network to capture more high-level semantic features that describe liquid impact events.

Residual Stage 4 (Layer 4) consists of two residual blocks. The first block performs yet another $2\times$ downsampling, increasing the channel dimensions from 256 to 512 and reducing the feature map size from $(256, 18)$ to $(512, 9)$. The second block maintains the same dimensions. After four stages of downsampling and successive channel expansion, the time-frequency representation is converted into a feature space with higher levels of abstraction, greater discriminative power, and improved generalization.

At last, an Adaptive Average Pooling (AdaptiveAvgPool2d) layer reduces the spatial dimensions of the 512-channel feature map ($512 \times 2 \times 9$) down to a single vector of size 512. This allows the retention of channel-wise information while eliminating spatial variation. Subsequently, a fully connected layer transforms this 512-dimensional feature vector into 2 dimensions, corresponding to the predicted probabilities of the non-water and water-intrusion states, respectively. In contrast to a flattening operation, global average pooling reduces the number of parameters significantly and enhances robustness to minor spatial shifts in the input feature maps.

The total architecture of the network is shown in Fig. 6.

At the core of ResNet architecture is the residual block. Each residual block consists of a processing chain beginning with a 3×3 convolutional layer (using a stride of 1 or 2 depending on the stage) followed by batch normalization and a ReLU activation. This is followed by a second 3×3 convolutional layer (with a fixed stride of 1) and one more batch normalization layer. Then, the output from this convolutional path is added to the original input element-wise (via a skip connection) and then a ReLU activation is applied to get the final output.

In cases where there is a mismatch between the number of input and output channels or if the stride of the first convolution is set to 2, there is a 1×1 convolutional projection layer that is added along the skip connection path to adjust both the dimensionality (and) and the spatial resolution of the input. This adjustment ensures that the input and convolutional output share the same dimensions, which validates the element-wise addition. It helps the network to learn a residual mapping rather than a direct mapping which helps with the vanishing gradient problem in deep networks and makes training more stable. The structure of the residual block is shown in Fig. 7.

As illustrated in Fig. 8, the feature maps of each layer in the ResNet-18 model are visualized. The shallow-layer features are relatively dispersed, reflecting low-level time-frequency representations. In the middle layers, concentrated yellow high-activation regions emerge, indicating that the network has learned to identify key discriminative feature locations. In the deep layers, the feature maps become sparse but exhibit strong activations, reflecting highly abstract and discriminative high-level features learned by the network. These observations clearly demonstrate the advantage of ResNet in progressively learning hierarchical feature representations, where the network extracts increasingly complex and task-relevant features layer by layer.

V. SIMULATION EXPERIMENTS

A. DATASET

Acoustic impact signals from cable elbow terminals were collected in a controlled laboratory setting. To ensure diversity and representativeness of the collected samples, we obtained a total of 1450 PCM-format audio recordings spanning six channels. The dataset was allocated into training and validation sets in a 7:3 ratio. All samples were preprocessed using Mel-spectrogram and transformed to a fixed 64×281 time-frequency representation (sampling rate $=$ 22,050 Hz; FFT window size $=$ 1024; frequency range $=$ 200–8000 Hz). This uniform representation helps to establish a solid foundation for deep learning analysis.

B. EXPERIMENTAL CONFIGURATION

Experiments were conducted using a machine running Windows 11 and featuring 16 GB of RAM. It was powered by an AMD Ryzen 7 9700X processor and an NVIDIA GeForce RTX 5060 GPU with 12 GB of memory, providing sufficient computing resources for model training and inference. Signal preprocessing, model implementation and evaluation of experiments were all performed using MATLAB 2024a.

C. MODEL EVALUATION METRICS

In this study, the classification task model performance is evaluated holistically using the most common and representative metrics and assessed from multiple angles, recognition ability from multiple angles. Recognition ability is assessed from multiple angles using these metrics: Accuracy (ACC) [19], Precision (P) [20], Recall (R) [21], Area Under the

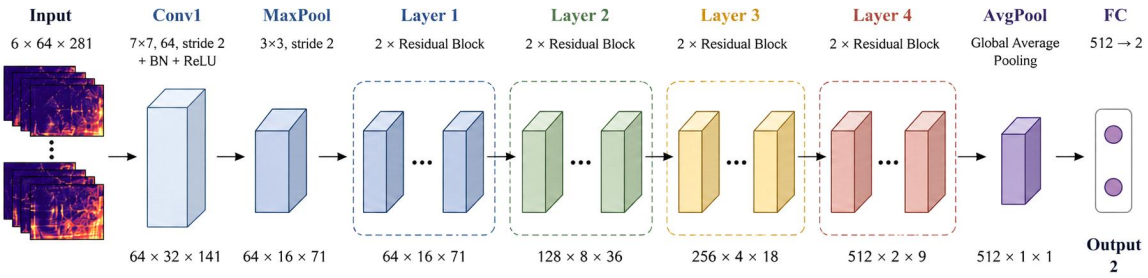

FIGURE 6. Structure of the ResNet.

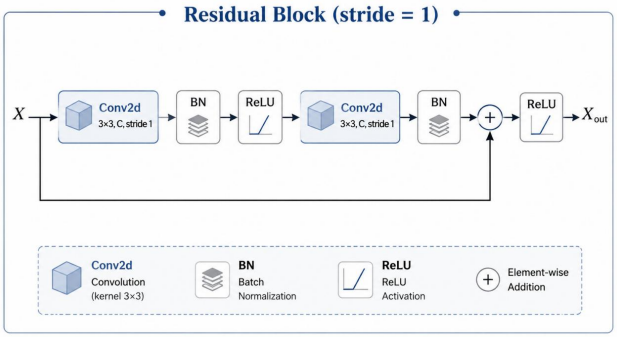

FIGURE 7(a). Structure of the One-Dimensional Convolutional Neural Network.

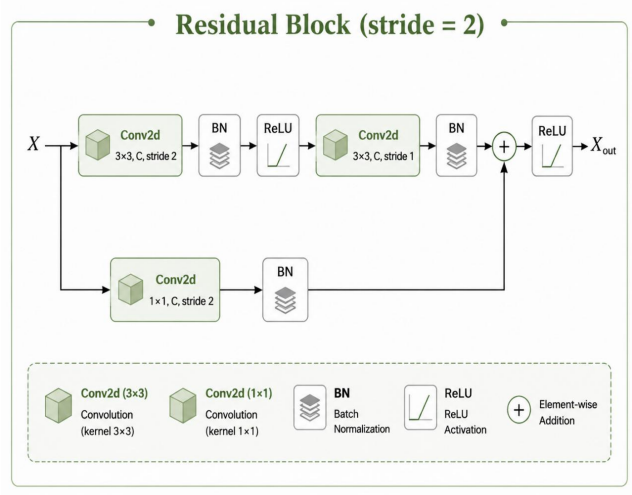

FIGURE 7(b). Structure of the Residual Block.

Receiver Operating Characteristic Curve (AUC) [22], and F1-score [23]. More specifically, precision is the measure of the false positive samples predicted as positive by the model out of the total positive predictions, and recall is the measure of correctly identified true positive samples among all true positive samples. The F1 score, or the harmonic mean of recall and precision, is a more informative score and summarizes the concern of trade-off for the two and is more informative when there is a concern for both false negatives and false positive detections. Here are these evaluation metrics:

$$\mathrm{ACC} = \frac{TP + TN}{TP + TN + FP + FN} \tag{18}$$

$$P = \frac{TP}{TP + FP} \tag{19}$$

$$R = \frac{TP}{TP + FN} \tag{20}$$

$$TPR = \frac{TP}{TP + FN} \tag{21}$$

$$FPR = \frac{FP}{FP + TN} \tag{22}$$

$$AUC = \int_0^1 TPR(FPR)\, d(FPR) \tag{23}$$

$$F1 = \frac{2 * P * R}{P + R} \tag{24}$$

where TP represents true positive samples; TN represents true negative samples; FP represents false positive samples; FN represents false negative samples; TPR represents true positive rate; FPR represents false positive rate.

D. PERFORMANCE COMPARISON BEFORE AND AFTER PREPROCESSING

To verify the impact of data augmentation strategies on the CNN model, three configurations are evaluated: a baseline CNN model (CNN-Baseline), a CNN model with speed perturbation (CNN+SP), and a CNN model with spectrogram masking (CNN+SM). The experimental results are shown in Fig. 9.

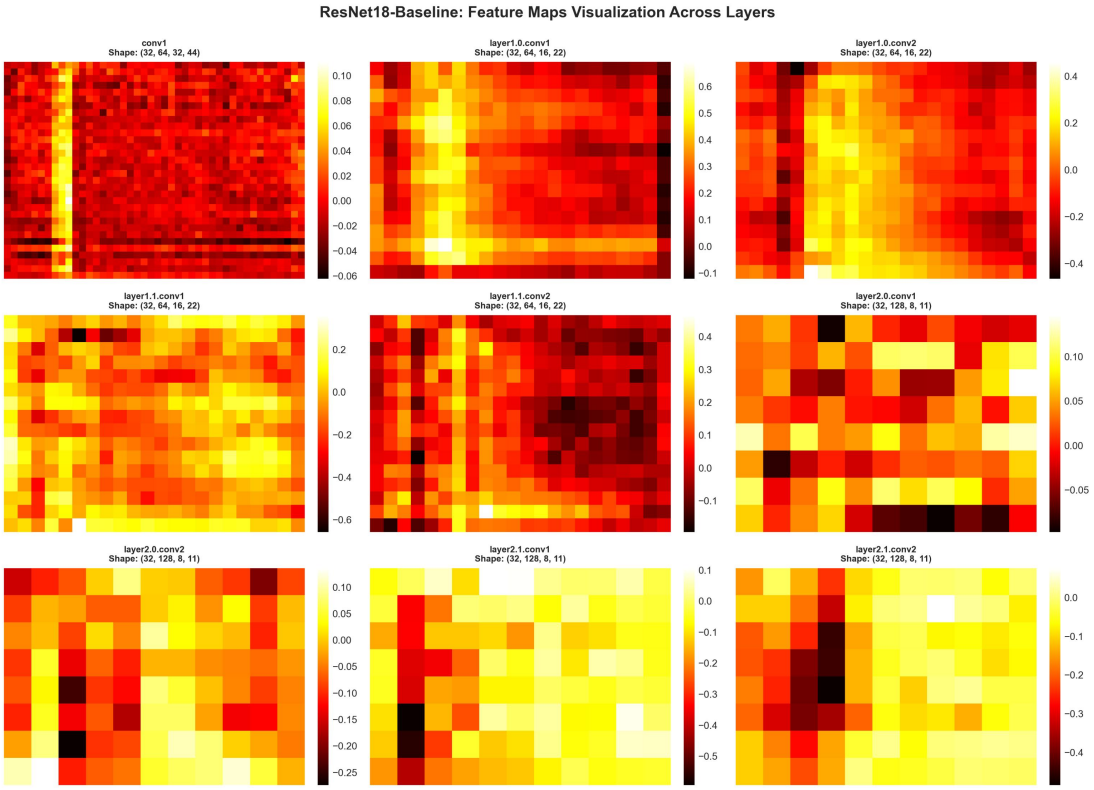

FIGURE 8. Feature Maps Across Layers of ResNet-18.

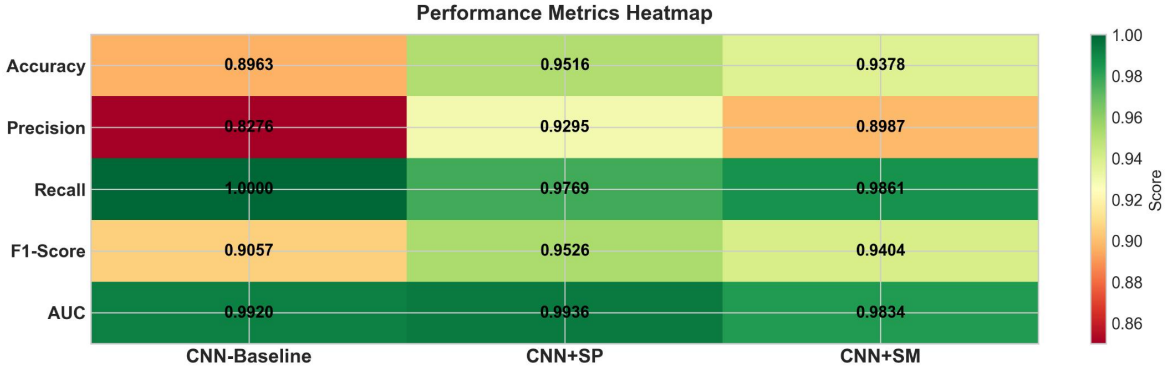

FIGURE 9. Comparison of CNN Classification Performance Before and After Data Augmentation.

The effectiveness of the proposed data augmentation methods is fully validated. Compared with the CNN-Baseline model, speed perturbation (SP) improves the accuracy from 89.63% to 95.16% and increases the F1-score from 89.57% to 95.26%. Similarly, spectrogram masking (SM) achieves an accuracy of 93.78% and an F1-score of 94.04%. Both augmentation strategies maintain a very high recall rate (97.69% and 98.61%, respectively) while significantly improving overall classification performance. These results demonstrate that introducing data augmentation in both the time and frequency domains can effectively and comprehensively enhance the liquid level detection performance of the CNN model.

E. PERFORMANCE COMPARISON BETWEEN CNN AND RESNET-18

To validate the advantages that deep residual networks have over shallow convolutional networks in the liquid level detection task, an architecture comparison experiment was carried out. Two models were trained under the same data preprocessing pipeline, training strategy, and hyperparameter settings: one, a CNN-Baseline model with three convolutional layers (roughly 0.2 million parameters); and two, a ResNet-18-Baseline model with eighteen convolutional layers (about 11.2 million parameters). By logging the full training process and the final validation performance, three key findings were obtained—the comparison of training

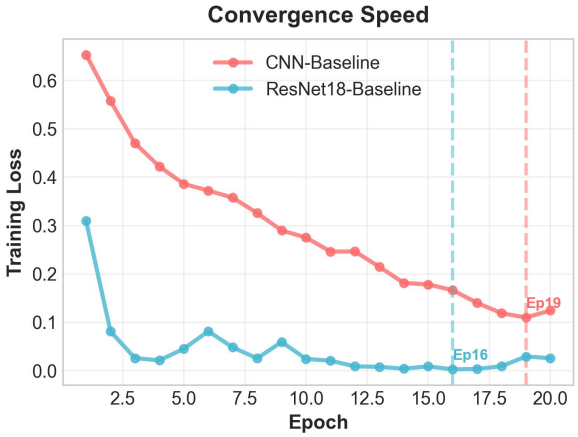

FIGURE 10. Comparison of Training Loss Curve Convergence Dynamics.

loss convergence dynamics (Fig. 10), along with the confusion matrices of the CNN-Baseline and ResNet-18-Baseline models (both shown in Fig. 11).

The two models follow quite different optimization paths. The CNN-Baseline training loss drops quickly over the first five epochs, after which the decline slows down, and it eventually bottoms out at a minimum loss of 0.12 around the 19th epoch. The ResNet-18-Baseline, on the other hand, produces a much smoother loss curve and converges in a faster and more stable manner, reaching a considerably lower loss of 0.03 by the 16th epoch. This is mostly due to the skip connections present in the residual learning, which offer efficient gradient pathways to make sure the earlier layers are not bypassed—even at the greater depths of the architecture.

With regard to classification performance, Fig. 11 shows how the CNN-Baseline model's confusion matrix is highly skewed. The accuracy of recognition of the non-water class is 80.73 percent and the false positive rate is 19.27 percent. In contrast, the water intrusion class is 100 percent accurate in classification. Such imbalance suggests that the model is biased towards and overfitting on one class, which is not ideal in real-world scenarios. ResNet-18-Baseline, on the other hand, is much more balanced across the two classes. The non-water class has a recognition accuracy of 95.87 percent (4.13 percent false positive rate) and the water intrusion class has an accuracy of 97.22 percent (2.78 percent false negative rate). The error rates in both classes are below 5 percent and the per class accuracies are above 95 percent. ResNet-18 has a total error rate of 3.46 percent, which is 6.22 percentage points lower than the 9.68 percent error rate for CNN.

In summary, ResNet-18 is superior to CNN-Baseline in accuracy, error rate, class balance, and stability of convergence, making it more suitable for applications in the detection of water ingress in cable elbows.

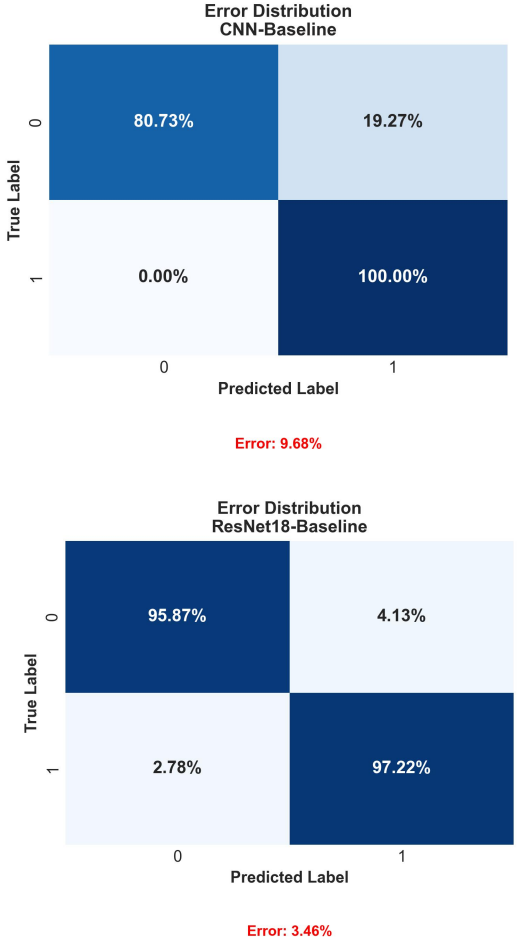

FIGURE 11. Confusion Matrix Comparison between CNN-Baseline and ResNet-18-Baseline.

F. OPTIMAL CONFIGURATION ANALYSIS OF RESNET-18

To examine whether ResNet-18 is the optimal configuration for the cable elbow joint water ingress acoustic fingerprint recognition task, I will be considering four common training configurations: The Baseline (no data augmentation), the Speed Perturbation Baseline (+SP), the Spectrogram Masking Baseline (+SM), and the combination of both augmentations in a single configuration (+SP+SM). I will represent the various F1-score results across different Signal-to-Noise Ratios (SNR) in Figure 12.

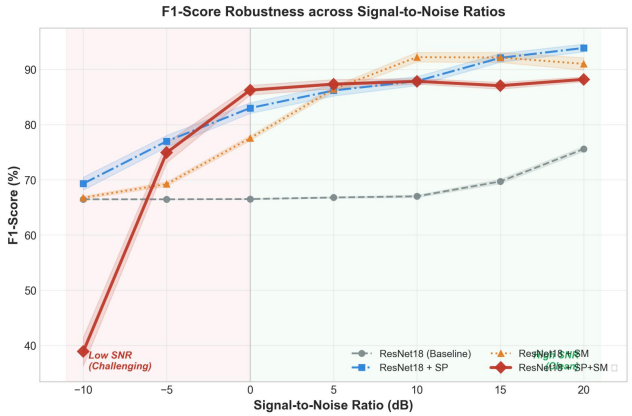

FIGURE 12. F1-Score Performance of Four Configurations under Different SNR Levels.

By observing the three different noise intervals, we can conclude that the best augmentation technique changes based on the gap of noise present. In the range of low SNR (0–5 dB), the combination of SP and SM yields the best results. SM covers a variety of methods including time masking, frequency masking, and combined variants of masking. This approach requires the model to learn how to utilize the most important residual features from the severely corrupted spectrograms. On the other hand, SP improves the model's ability to handle changes in the temporal domain that are of varying lengths (e.g. varying lengths of silence in the audio stream). The combination of these two methods results in a significant improvement in the model's performance when noise levels are high.

Spectral information appears to retain some discriminative power, while the additional temporal perturbations introduced by SP may inhibit the learning of frequency-domain features, resulting in an over-augmentation effect. Thus, removing SP while retaining SM is preferred.

Once we shift our focus to the high-SNR domain, the spectral information is sufficiently clean, and the SM module begins to add extraneous random masking that removes critical discriminative detail. Here, SP alone becomes the preferred option, as it maintains the integrity of the spectral envelope while enhancing temporal robustness to varying rhythmic impacts.

Thus, the optimal configuration for a given ResNet-based model should be selected in accordance with the prevailing environmental SNR.

VI. CONCLUSION

This paper has presented an acoustic fingerprint recognition method integrating data augmentation with deep residual networks for detecting concealed water accumulation in cable elbow terminations. The main findings and contributions are summarized as follows:

(1) Both Speed Perturbation (SP) and Spectrogram Masking (SM) significantly enhance recognition performance. SP improves CNN accuracy from 89.63% to 95.16%, while SM achieves 93.78% accuracy, demonstrating the effectiveness of dual-domain data augmentation.

(2) ResNet-18 reduces the total error rate from 9.68% (CNN baseline) to 3.46%, achieving balanced class-wise accuracy above 95% for both water and non-water states, with improved convergence stability.

(3) The optimal augmentation strategy depends on the SNR level: SP+SM performs best in low-SNR environments (0–5 dB), SM alone is optimal for moderate-SNR conditions (5–15 dB), and SP alone suffices for high-SNR scenarios (> 15 dB). This finding provides practical guidance for field deployment.

The proposed method offers notable practical advantages, including non-contact implementation, easy deployment, and high detection accuracy ($> 95\%$), making it a promising solution for intelligent condition monitoring of power distribution equipment. However, the current study is limited to laboratory validation and binary classification. Future work will focus on field deployment trials, extension to multi-level moisture quantification, and integration with complementary sensing modalities for comprehensive diagnosis.

ACKNOWLEDGEMENTS

This work was supported by the Science and Technology Project of State Grid Corporation of China (Grant No. SGAHXC00JTJS2500623)

REFERENCES

[1] H. Yang, J. Wang, P. Zhao, C. Yu, H. You, and J. Dou, "On-Line Insulation Monitoring Method of Substation Power Cable Based on Distributed Current Principal Component Analysis," Energies, vol. 18, no. 3, pp. 20, 2025. DOI: 10.3390/en18030688.

[2] Y. Li, L. Jiang, M. Xie, J. Yu, L. Qian, K. Xu, M. Chen, and Y. Wang, "Advancements and Challenges in Power Cable Laying," Energies, vol. 17, no. 12, pp. 26, 2024. DOI: 10.3390/en17122905.

[3] J. Tao, S. U. Rehman, R. Ali, and S. A. Raza, "Advancement and challenges: A review of power cable aging monitoring and diagnostic techniques," Renew. Sust. Energ. Rev., vol. 222, pp. 20, 2025. DOI: 10.1016/j.rser.2025.115970.

[4] P. Meng, T. Li, K. Zhou, Z. Tang, G. Zhu, Z. Li, Y. Cao, H. Zhang, Y. Yin, and J. Guo, "A Novel Time-Frequency-Domain Reflectometry Location Method for Power Cable Defects Based on Synchrosqueezing Transform," IEEE Trans. Instrum. Meas., vol. 73, pp. 9, 2024. DOI: 10.1109/TIM.2024.3418110.

[5] C. C. Uydur and O. Arikan, "Dielectric performance analysis of laboratory aged power cable under harmonic voltages," Electr. Eng., vol. 104, no. 6, pp. 4197–4212, 2022. DOI: 10.1007/s00202-022-01614-4.

[6] G. Zhu, Z. Liu, K. Zhou, Z. Wang, L. Lu, Y. Li, and P. Meng, "A Novel Dampness Diagnosis Method for Distribution Power Cables Based on Time-Frequency Domain Conversion," IEEE Trans. Instrum. Meas., vol. 71, pp. 9, 2022. DOI: 10.1109/TIM.2022.3160840.

[7] G. Zhu, Z. Liu, S. Pan, P. Meng, K. Zhou, X. Wang, Q. Shen, and H. Zhou, "A Dampness Discrimination Method for MV Power Cable Joints Based on PDC Testing Under Thermal Excitation Conditions," IEEE Trns. Dielectr. Electr. Insul., vol. 32, no. 1, pp. 581–588, 2025. DOI: 10.1109/TDEI.2024.3487957.

[8] L. Wang, B. Wang, H. Ma, J. Zhang, Y. He, H. Wang, and H. Zhang, "Power Cable Status Evaluation Method Based on Electrical Tree Growth and Data Association Rules," Processes, vol. 12, no. 9, pp. 1, 2024. DOI: 10.3390/pr12092026.

[9] T. G. Bade, J. Roudet, A. Derbey, A. de Andrade, and L. Sadi-Haddad, "Adapter for the Impedance Measurement of Power Cable With an Impedance Analyzer," IEEE Trans. Electromagn. Compat., vol. 65, no. 3, pp. 705–715, 2023. DOI: 10.1109/TEMC.2023.3250107.

[10] G. Li, J. Chen, H. Li, L. Hu, W. Zhou, C. Zhou, and M. Li, "Diagnosis and Location of Power Cable Faults Based on Characteristic Frequencies of Impedance Spectroscopy," Energies, vol. 15, no. 15, pp. 18, 2022. DOI: 10.3390/en15155617.

[11] A. Ghaforian, P. Duggan, and L. Lu, "A Comprehensive Review of Cable Monitoring Techniques for Nuclear Power Plants," Energies, vol. 18, no. 9, pp. 26, 2025. DOI: 10.3390/en18092333.

[12] C. S. Subudhi and S. K. Sen, "Non-Intrusive Online Time Domain Reflectometry Technique for Power Cables," IEEE Trans. Power Deliv., vol. 39, no. 1, pp. 101–110, 2024. DOI: 10.1109/TPWRD.2023.3324655.

[13] J. Wang, Z. Zhang, R. Zhou, A. Fan, T. Zhai, T. Guo, and Y. Zhao, "A Novel Fiber-Optically Powered Partial Discharge Monitoring System for Cable Joints Based on Acoustic Emission," IEEE Sens. J., vol. 25, no. 20, pp. 38809–38819, 2025. DOI: 10.1109/JSEN.2025.3605966.

[14] D. F. J. Jabha, R. Joselin, and R. Sowmya, "Examining the Use of Acoustic Emission Technique for Evaluating Partial Discharge in Power Cables: A Review," J. Fail. Anal. Prev., vol. 24, no. 5, pp. 2316–2326, 2024. DOI: 10.1007/s11668-024-02006-5.

[15] W. Tang, K. Brown, D. Mitchell, J. Blanche, and D. Flynn, "Subsea Power Cable Health Management Using Machine Learning Analysis of Low-Frequency Wide-Band Sonar Data," Energies, vol. 16, no. 17, pp. 17, 2023. DOI: 10.3390/en16176172.

[16] J. Lee, J. Park, and H. Park, "Optimizing Cable Replacement Schedules in Power Grids: A Data-Driven Approach with Machine Learning and Power Flow Analysis," IEEJ Trans. Electr. Electron. Eng., vol. 20, no. 9, pp. 1497–1502, 2025. DOI: 10.1002/tee.70056.

[17] H.-W. Sian, C.-C. Kuo, S.-D. Lu, and M.-H. Wang, "A novel fault diagnosis method of power cable based on convolutional probabilistic neural network with discrete wavelet transform and symmetrized dot pattern," IET Sci. Meas. Technol., vol. 17, no. 2, pp. 58–70, 2023. DOI: 10.1049/smt2.12130.

[18] M.-H. Wang, S.-D. Lu, and R.-M. Liao, "Fault Diagnosis for Power Cables Based on Convolutional Neural Network With Chaotic System and Discrete Wavelet Transform," IEEE Trans. Power Deliv., vol. 37, no. 1, pp. 582–590, 2022. DOI: 10.1109/TPWRD.2021.3065342.

[19] J. Zhao, X. Han, Z. Niu, G. Zhang, J. Yang, B. Li, and Y. Wu, "Power System Operation State Identification Considering the Process of Transient Stability," Proceedings of the Chinese Society of Electrical Engineering, vol. 44, no. 20, pp. 7970–7982, 2024. DOI: 10.13334/j.0258-8013.pcsee.231219.

[20] L. Xi, D. Peng, W. Cao, H. Chen, F. Bai, and W. Wang, "FDIA Location Detection for Data-driven Algorithms in Cyber-physical Power Systems," Proceedings of the Chinese Society of Electrical Engineering, vol. 45, no. 18, pp. 7110–7122, 2025. DOI: 10.13334/j.0258-8013.pcsee.240412.

[21] B. Zhao, X. Liu, L. Zhang, Z. Chen, L. Zhang, and C. Xie, "Fault Diagnosis of Proton Exchange Membrane Fuel Cell Integrated System Based on GoogleNet and Transfer Learning," Proceedings of the Chinese Society of Electrical Engineering, vol. 44, no. 13, pp. 5147–5157, 2024. DOI: 10.13334/j.0258-8013.pcsee.230210.

[22] L. Gao, D. Wang, L. Zhuang, X. Sun, M. Huang, and A. Plaza, "BS^{3}LNet: A New Blind-Spot Self-Supervised Learning Network for Hyperspectral Anomaly Detection," IEEE Trans. Geosci. Remote Sensing, vol. 61, pp. 18, 2023. DOI: 10.1109/TGRS.2023.3246565.

[23] J. Ircio, A. Lojo, U. Mori, S. Malinowski, and J. A. Lozano, "Minimum Recall-Based Loss Function for Imbalanced Time Series Classification," IEEE Trans. Knowl. Data Eng., vol. 35, no. 10, pp. 10024–10034, 2023. DOI: 10.1109/TKDE.2023.3268994.