IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 31 January 2024, accepted 14 February 2024, date of publication 16 February 2024, date of current version 23 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3366807

RESEARCH ARTICLE

Acoustic-Based Online Monitoring of Cooling Fan Malfunction in Air-Forced Transformers Using Learning Techniques

REZA NEMATIRAD[1], (Graduate Student Member, IEEE), MEHDI BEHRANG[2], AND ANIL PAHWA[1], (Life Fellow, IEEE)

[1]Department of Electrical and Computer Engineering, Kansas State University, Manhattan, KS 66506, USA
[2]Department of Electrical and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada

Corresponding author: Reza Nematirad (nematirad@ksu.edu)

This work was supported by Kansas State University Open Access Publishing Fund.

ABSTRACT Cooling fans are one of the critical components of air-forced dry-type transformers for regulating internal temperatures. Therefore, effective malfunction detection is crucial to maintain the transformer temperature within an acceptable range and prevent overheating. Current malfunction detection of cooling fans in certain types of transformers relies on complementary indicators, such as top-oil temperature, oil convection, dissolved gas, and oil quality. However, these conventional indicators are not directly applicable to air-forced transformers, which primarily use cooling fans as their cooling system. To overcome this challenge, this study utilizes cooling fan audio records as indicators. The audio signals are classified into normal and malfunctioning classes using advanced learning algorithms, including convolutional neural networks and random forests. Learning algorithms require transforming recorded audio data into proper formats. Accordingly, convolutional neural networks are trained based on spectrogram images derived from audio signals. For random forests, various time-frequency feature extraction methods are used to derive meaningful representations from audio signals. Besides, multiple data augmentation techniques are employed to enhance the dataset size and diversity. Algorithmic performance is optimized through hyperparameter tuning and classifier threshold adjustment. To further validate the model, a test is conducted on another dataset to evaluate the fitted learning model applicability in real-world applications. Simulations reveal that convolutional neural networks outperform random forests, whereas the latter provides superior interpretability of acoustic features compared to the former.

INDEX TERMS Acoustic analysis, air-forced power transformers, cooling fans, data augmentation, fault detection, joint time-frequency feature extraction, learning algorithms.

I. INTRODUCTION

Transformers play a vital role in electrical power systems by stepping up or down voltage. Transformers can be categorized into different types based on their specific functions, including distribution, power, and instrument transformers [1]. Regardless of the transformer type, the reliable operation of transformers is crucial to ensure the stability and quality of the power supply. The copper losses result from the electrical resistance in the transformer winding, and iron losses associated with the transformer magnetic core increase the temperature of the windings and the core [2]. Exposing transformers to excessive heat may lead to insulation degradation, component decomposition, reduced lifespan, and even catastrophic failure [3]. To prevent the overheating of transformers, cooling systems are essential. By transferring excess heat to an external medium, a cooling system efficiently dissipates heat generated within transformers. A proper cooling system not only enhances the lifespan of transformers but also contributes to energy efficiency, reducing the

The associate editor coordinating the review of this manuscript and approving it for publication was Gerard-Andre Capolino.

operational costs associated with transformer maintenance [4]. The design of transformer cooling systems varies based on the transformer type. These systems may incorporate specific components, such as oil, pumps, fans, and radiators to regulate and maintain transformers within specified temperature limits.

A. TRANSFORMER COOLING SYSTEM METHODS

Based on the cooling medium employed, transformers are commonly classified into two primary categories: dry-type transformers, which rely on air, and oil-type transformers, which utilize oil-air [5].

1) DRY-TYPE TRANSFORMERS

The dry-type transformers dissipate heat generated during operation by using air as a cooling medium. Cooling methods for dry-type transformers typically are based on air-natural (AN) and air-forced (AF) methods. The AN method is generally utilized in small transformers (less than 3 MVA) and uses natural air circulation for cooling [6], [7]. However, natural air is insufficient to cool the transformer with higher loads. In this case, additional fans are embedded to improve airflow and transformer cooling system efficiency as the AF method [7].

2) OIL-TYPE TRANSFORMERS

The oil-type or oil-immersed transformers use oil-air as a cooling medium. Oil circulates through windings and core transferring heat away from the transformer. Additionally, oil provides insulation for transformers [8]. A variety of methods can cool oil-type transformers. The oil-natural air-natural (ONAN) method dissipates heated oil naturally into the surrounding air without relying on forced cooling mechanisms. In contrast, the oil-natural air-forced (ONAF) method uses external fans. In the oil-forced air-forced (OFAF) method, external pumps and fans facilitate oil and air circulation [9].

B. MONITORING OF COOLING SYSTEMS

Monitoring and detecting malfunctions in cooling systems is critical for preventing transformer overheating. Various studies have been conducted to monitor the cooling system conditions of transformers. Offline and online methods are typically used to assess the condition of the cooling system of transformers.

1) OFFLINE AND ONLINE MONITORING

Offline methods refer to regular maintenance checks conducted periodically outside the real-time operation of the transformer. However, the cooling systems may malfunction between maintenance periods, resulting in overheating and potential problems. Moreover, these inspections can be costly, requiring disconnecting transformers from the grid [10]. On the other hand, online methods can continuously monitor cooling fan conditions. Most utilities only have access to online information on whether the cooling fans are working or not. However, this method cannot detect cooling systems malfunctioning [11]. Despite their limitations, online methods have the potential for malfunctioning detection allowing for the timely detection of faults and early alerts [12].

2) OVERVIEW OF ONLINE COOLING SYSTEM MONITORING TECHNIQUES

Various online cooling system monitoring techniques have been employed across transformer types. For example, in [13], a real-time monitoring system using the Internet of Things and the global system for mobile communications is presented for oil-immersed transformers, focusing on parameters like oil temperature, winding temperature, and relay status. The thermal modeling approach outlined in reference [14] employed ambient, top-oil, winding, and radiator temperatures as indicators for estimating the top-oil and radiator temperatures in oil-immersed AF transformers. The authors of [15] utilized differential pressure sensors for thermal monitoring conditions in oil-immersed transformers by comparing the humidity of incoming and outgoing air and assessing the behavior of On-line tap changers shape. A novel approach for monitoring oil-immersed transformers by integrating various monitoring techniques such as thermal, vibrational, and dissolved gas analysis was proposed in [16]. This integration allows for enhanced predictive analytics and efficient failure prevention.

In the existing literature, malfunction detection of cooling fans, which serve as critical components in AF transformers and auxiliary equipment in oil-type transformers, has not received much attention. A novel approach to early fault detection in oil-immersed transformer fans was proposed in [11]. Based on existing top-oil temperature data, the oil exponent concept was utilized as a parameter of fan operation. A physical model for malfunction detection in cooling fans of ONAF power transformers was presented in [11] by monitoring top-oil temperature and using an oil exponent as the critical criterion. In [17] an improved online presented method was proposed for estimating the hot-spot temperature of ONAN transformers when auxiliary cooling fans are employed. A novel online algorithm for fan malfunction detection in ONAF transformers integrated with renewable energy resources was proposed in [18]. This algorithm relies on detecting changes in the estimated parameters of top-oil temperature, such as oil convection, ambient temperature, and load factor.

3) AF TRANSFORMER COOLING FAN MONITORING

While various studies were conducted on the malfunction detection of cooling fans of transformers, a particular gap still remains. Many existing studies focused on online fault detection of oil-type transformer cooling fans. In such studies, the malfunctioning detection of cooling fans relies on complementary indicators, such as top-oil temperature, oil convection, dissolved gas analysis, and oil quality. However,

such indicators are not applicable to AF transformers, which primarily use cooling fans as their cooling system. Therefore, an independent online monitoring method is essential for the malfunctioning detection of AF transformers ensuring reliable operation, preventing overheating, and extending the transformer lifetime.

Recognizing this gap in the existing literature on cooling fan malfunction of AF transformers, this study presents an innovative acoustic-based method that utilizes audio signals from AF transformers as an effective alternative indicator. Acoustic data is non-intrusive, allowing microphones to be placed near cooling fans without requiring significant modifications to existing infrastructure. It also enables real-time monitoring, providing instant alerts and minimizing latency in malfunction detection. Additionally, acoustic data collection is cost-effective and easy to install [19].

A simple microphone was used to collect the required data for the learning algorithms. Various data augmentation techniques were employed to increase the diversity and number of the training dataset. Random forests (RFs), a machine learning algorithm, and convolutional neural networks (CNNs), a deep learning model, were utilized to train the online models. CNNs were trained on spectrogram images derived from audio signals. This allows CNNs to learn relevant features from the spectrogram representations automatically, whereas RFs require a distinct process to extract the features from audio signals. Accordingly, different joint time-frequency feature extraction approaches were used to capture relevant acoustic characteristics from the raw audio signals. Then, a classification task was conducted to group the audio signals into normal and malfunctioning classes by using the RFs and CNNs. Hyperparameter tuning and classifier threshold optimization were utilized to enhance the learning algorithm performances. Finally, the trained learning models were evaluated with different metrics. It should be noted that the methods proposed in this study can be applied to other types of transformers that are equipped with cooling fans.

C. CONTRIBUTIONS

The main contributions of this study can be summarized as follows:

- Introduction of the machine and deep learning for malfunction detection: This study develops machine learning and deep learning techniques in AF transformer cooling fan malfunction detection. This approach addresses a specific gap in the literature, offering a more sophisticated and accurate method than traditional monitoring techniques.
- Innovative use of acoustic data for detection: This study introduces alternative indicators for monitoring transformer cooling fans by employing simple microphones to gather acoustic data. This is essential since traditional indicators, such as top-oil temperature, oil convection, dissolved gas analysis, and oil quality, do not apply to air-forced transformers. This approach offers several key advantages, including the ability to online malfunction detection for immediate corrective actions, which is crucial for maintaining transformer efficiency and safety. Furthermore, this method is cost-effective compared to more complex monitoring indicators and is non-intrusive and easy to install without requiring significant modifications to existing infrastructure.
- Advanced optimization and validation of learning models: Recognizing that the default classifier threshold of 0.5 may not be optimal for the learning models, we conducted a thorough optimization process to identify the most effective classifier threshold. Besides, Bayesian optimization is utilized to fine-tune the hyperparameters of the learning models. To validate the performance of the optimized models, we employed learning curves and varied the training sample sizes. This systematic approach allowed us to comprehensively assess the capabilities of learning models and their robustness against different data quantities. To confirm the potential reliability of the best-fitted model for real-world applications, a critical test was conducted using a different dataset obtained from another AF transformer in a distinct ambient environment.

II. METHODOLOGY

This section outlines the techniques employed for malfunction detection in AF transformer cooling fans.

A. DATA COLLECTION AND PREPROCESSING

This study collected a dataset of 300 recorded voices from an AF transformer cooling fan. The audio signals had a sample rate of 44100 per second and a length of 225792, resulting in an approximate duration of 5.12 seconds. These recordings were obtained from an AF transformer in a closed room away from an urban area. The dataset consists of 200 instances recorded during the normal fan operation and 100 instances captured during the fan malfunctions at different loads and fan speeds. The recording process involved using a high-quality microphone close to the transformer cooling fan to capture acoustic signals accurately. The data were organized with their corresponding labels and divided into three subsets, including training, validation, and testing. The training subset was used to train the learning models. The validation subset was used to evaluate the performance of the model during the training phase and tuning of hyperparameters. The test subset contained unseen data that the model was not exposed to them during the training or validation step. In addition, the optimal fitted learning model was tested with a new dataset comprising audio recordings from another AF transformer situated in a non-isolated room environment in an urban area. This setting introduces a diverse range of ambient noises and operational variations, providing a more rigorous testing ground for the model applicability. This new dataset included 20 normal and 15 malfunctioning recordings, representing real-world operational conditions.

1) DATA AUGMENTATION

Since the number of data samples was limited, the data augmentation techniques were used to artificially expand the size of the training dataset by creating additional samples as it helps mitigate the risk of overfitting and improves the generalization capability and robustness of the learning models [20]. It is essential to ensure that data augmentation does not distort the integrity of the original signals and to avoid applying data augmentation to validation and test datasets [21].

Data augmentation has a wide application in learning algorithms. In the context of deep learning, particularly in image processing, techniques such as flipping, rotation, image cropping, and shifting are widely utilized to diversify the training dataset. However, applying these traditional image-based augmentation techniques to spectrogram images derived from audio signals may substantially alter or distort their time-frequency characteristics. To address this challenge, the SpecAugment data augmentation techniques were developed, specifically designed for augmenting spectrograms without altering their acoustic characteristics [22]. The SpecAugment methods of frequency and time masking techniques that have been shown to increase the diversity of the spectrograms effectively were employed in this study [23]. A frequency mask sets a band of frequencies as $[f_0, f_0 + f)$ in the spectrogram to zero, where f is the consecutive frequency to be masked. Number of consecutive frequencies of f was set from the uniform distribution $f \sim U(0,\ F)$, where F is a frequency mask parameter. Besides, f_0 was adjusted from the interval $(0,\ v - f)$, where v is the number of frequency channels. This simulates the effect of removing a specific range of frequencies from the audio signals. As a result, the model becomes more robust to missing frequency bands when trained on augmented data. Time masking involves selecting a time segment and setting all values in the spectrogram for that segment to zero. The procedure for setting consecutive time steps to zero is similar to that for frequency masking, except instead of frequency, time is used. Consequently, the model is robust to occurrences in the real-world such as short silences or missing audio. The selection of frequency and time mask parameters are mainly arbitrary and per the recommendation of [23].

Furthermore, this study utilized a range of the most common audio signal augmentation methods, including time stretching, time shifting, and noise injection, to expand the training data sets.

Time stretching involves modifying the speed of audio signals while preserving their pitch. This method could simulate variations in fan speed. By randomly adjusting the playback speed from 0.8 to 1.5, we generated new audio samples representing the fan operating at different speeds [24].

Time shifting involves moving an audio signal forward or backward in time. This technique allows us to simulate changes in the start time of the fan operation. By randomly introducing time shifts to the audio recordings, we created additional samples reflecting different instances when the cooling fan began its operating [25].

Noise injection involves adding various types and levels of noise to the audio samples to simulate real-world recording conditions. In this study, we employed two distinct noise injection methods, such as white noise and environmental noise, each serving a specific purpose. White noise is a fundamental source of randomness, characterized by equal energy distribution across all frequencies and a flat power spectral density [26]. Using this method allows us to convey the unpredictability that is present in real-world audio recordings into our audio samples. To add white noise to the cooling fan audio signals, the process involves the following steps [27]:

1. Calculating the power of the original audio signal:
$$P_{signal} = \frac{1}{N}\sum_{i=}^{N} |x_i(t)|^2 \tag{1}$$
where N is the length of the original audio signal $x_i(t)$ and P_{signal} represents the power of the original audio signal, respectively.
2. Select a random Signal-to-Noise Ratio (SNR) within a range from 7 to 20 dB [28].
3. Calculating of the desired noise power:
$$P_{noise} = \frac{P_{signal}}{10^{(\frac{SNR}{10})}} \tag{2}$$
where P_{noise} indicates the desired noise power.
4. Generating white noise $\omega_i(t)$ by sampling from a normal distribution with mean 0 and standard deviation of 1.
5. Scaling the white noise to match the desired noise power:
$$SF = \sqrt{\frac{P_{noise}}{\sigma^2}} \tag{3}$$
$$\omega_i^s(t) = SF \times \omega_i(t) \tag{4}$$
where SF, σ^2, and $\omega_i^s(t)$ indicates scale factor, variance of the white noise, and scaled white noise, respectively.
6. Adding the scaled white noise to the original audio signal:
$$y_i(t) = \omega_i^s(t) + x_i(t) \tag{5}$$

where $y_i(t)$ is the injected audio signal with the white noise.

Environmental noise includes various background sounds and disturbances typically present in a specific setting. This study incorporated a range of environmental noise sources, including thunder, rain, wind, and ambient noise. The processes for adding environmental noise to cooling fan audio signals are the same as for adding white noise. To simulate the conditions more realistically, we have used a range of different SNRs from 1 to 40 dB to account for variability and fluctuation in environmental noises [29].

Audio augmentation techniques can be applied parallelly and sequentially. In parallel augmentation, multiple augmentation techniques are applied simultaneously to the audio

data. It is particularly useful to generate a diverse set of augmented data. Whereas sequential augmentation is applied one after another. This approach can be beneficial when there is a need to expand a limited dataset by generating additional samples. In this study, both parallel and sequential data augmentation approaches were utilized to take advantage of the benefits associated with them.

B. FEATURE EXTRACTION AND SPECTROGRAM REPRESENTATION

Since the recorded audio data were in the form of signals and thus unsuitable for learning algorithms, they needed to be converted into an appropriate format. Accordingly, feature extraction techniques were utilized in RFs and CNNs leverage spectrogram representation.

1) FEATURE EXTRACTION TECHNIQUES

Feature extraction is a crucial step for detecting malfunctions from the signals because the training and testing of RF algorithms highly depend on the features used to develop them [30]. This is achieved through feature extraction, which transforms the audio signals into representative features. This study employed various joint time-frequency feature extraction methods to derive meaningful representations from audio signals. Audio signals were divided into smaller windows or segments to calculate time-frequency domain features from a signal. For each window, the time and frequency features were calculated. Besides, the following window should overlap with the previous window to prevent loss of information at the window edges. Choosing the window length and overlapping time are mainly arbitrary and per the recommendation of [31].

We employed the following time-domain feature extraction methods that are effective in identifying the dominant features of the audio signals [32]:

- RMS quantifies the amplitude of a signal over a certain time period and represents the overall energy of the signal as follows:

$$\text{RMS} = \sqrt{\frac{1}{K}\sum_{i=1}^{K} x_i^2} \tag{6}$$

where K is the number of samples in a window.

- ZCR counts how many times the signal crosses the zero-amplitude line as follows:

$$\text{ZCR} = \frac{1}{K-1}\sum_{i=1}^{K-1} |\text{sing}(x_i) - \text{sing}(x_{i-1})| \tag{7}$$

- Kurtosis factor quantifies the extent to which a distribution is heavy-tailed or light-tailed relative to a normal distribution and could be expressed as follows:

$$\text{kurtosis factor} = \frac{\frac{1}{K}\sum_{i=1}^{K}(x_i - \frac{1}{K}\sum_{i=1}^{K} x_i)^4}{(\frac{1}{K}\sum_{i=1}^{K} x_i^2)^2} \tag{8}$$

A positive kurtosis suggests heavier tails, a negative kurtosis suggests lighter tails and a kurtosis of 0 indicates normal tails.

- The shape factor provides insight into the duration and relative proportions of the positive and negative peaks in a signal amplitude by describing its shape or waveform as follows:

$$\text{Shape factor} = \frac{\text{Duration of positive peaks}}{\text{Duration of negative peaks}} \tag{9}$$

- The crest factor measures the ratio between the peak amplitude and RMS as follows:

$$\text{Crest factor} = \frac{\text{Peak Amplitude}}{\text{RMS}} \tag{10}$$

- The impulse factor is used to characterize the impulsiveness or transient nature of a signal. It provides information about sudden changes or impulses within a signal and could be expressed as follows:

$$\text{Impulse Factor} = \frac{\sum_{i=1}^{K} x_i}{K \times \text{RMS}} \tag{11}$$

- Besides, statistical measures such as the mean, variance, minimum, and maximum of signals were calculated for each window.

To calculate the frequency-domain features, the signals were transformed from the time domain to the frequency domain using the Fast Fourier Transform (FFT) as follows [34]:

$$X_i(f) = FFT\{x_i(t)\} \tag{12}$$

where$X_i(f)$ is the frequency domain of $x_i(t)$.

After this transformation, the following frequency-domain feature extraction methods that have demonstrated great capability in capturing information about the energy distribution of audio signals were employed [33]:

- Spectral centroid calculates the location of the mass center of a spectrum providing an estimate of the dominant frequency in the signal and could be expressed as follows [34]:

$$\text{Spectral Centroid} = \frac{\sum_{i=1}^{K} f_i \times X_i}{\sum_{i=1}^{K} X_i} \tag{13}$$

where, f_i is the frequency of bin i in the FFT of the window.

- The Spectral bandwidth is the standard deviation of the distribution of spectral components around the spectral centroid as follows [35]:

$$\text{Spectral bandwidth} = \sqrt{\frac{\sum_{i=1}^{K}(f_i - \text{Spectral Centroid})^2 \times X_i}{\sum_{i=1}^{K} X_i}} \tag{14}$$

- Spectral flatness measures the uniformity of the power spectrums of a signal in the frequency distribution. Mathematically, spectral flatness is the ratio of the geometric mean to the arithmetic mean of the power spectrums [36]:

$$\text{Spectral flatness} = \frac{\sqrt[K]{\prod_{i=0}^{K-1} X_i}}{\frac{1}{K}\sum_{i=0}^{K-1} X_i} \tag{15}$$

- Spectral flux represents the rate of change in the spectral content of a signal providing information about how quickly the frequency energy distribution of the signals change. The spectral flux is the 2-norm of the difference between the magnitude spectra of consecutive frames [36].

$$\text{Spectral flux} = \left\| \sqrt{|X_i|} - \sqrt{|X_{i-1}|} \right\|_2 \tag{16}$$

- Peak frequency can be determined by identifying the frequency value associated with the maximum power in the signals [37].

Feature extraction was performed over short segments of the audio signals. Therefore, each feature extraction technique generated a vector of features where each value corresponds to a specific window in an audio signal. These vectors represent the time and frequency characteristics of an audio signal. It should be noted that using multiple and various feature extraction techniques helps capture various aspects and characteristics of the audio signal, improving the performance, robustness, and accuracy of the learning algorithms.

2) SPECTROGRAM REPRESENTATION TECHNIQUE

Image processing techniques have shown remarkable versatility, finding applications across various domains, including multimedia. For instance, recent advancements in nighttime image enhancement and haze removal [38], [39], [40], [41] demonstrate the adaptability of image processing methods in challenging visibility conditions. These developments underline the potential of image-based approaches [42], [43] in diverse settings, paving the way for their application in other fields, such as audio signal processing. In this study, we extend the concept of image processing to the realm of audio analysis. Inputs to CNN algorithms typically consist of multi-dimensional arrays of pixel values that encode visual information. Since CNN models have been extensively developed and trained on image datasets, using image representations allows us to leverage predefined deep learning models that have demonstrated success in various computer vision tasks [44]. To leverage the power of deep learning models for audio processing, a common practice is to convert audio signals into spectrogram image representations [45]. This technique can transform the temporal evolution of audio signals into a multi-dimensional visual representation. This transformation enables CNN models to extract meaningful audio features and learn complex patterns from spectrograms. Accordingly, the spectrogram technique was utilized in this study to develop the CNN classification model. Spectrograms were generated from audio signals as follows [46]:

1. Using a windowing technique, audio signals were divided into short overlapping segments.
2. FFT was applied to each segment to transform the audio signal from the time domain to the frequency domain.
3. The power spectrum $P_i(f)$ was calculated by taking the squared magnitude of the Fourier transform.

$$P_i(f) = |X_i(f)|^2 \tag{17}$$

4. The spectrograms are created by plotting the power spectrum values as a heat map or a grayscale image. The X-axis represents time, the Y-axis indicates frequency, and the intensity or color of each pixel indicates the magnitude or power of the corresponding frequency component.

One significant advantage of using spectrogram representations over feature extraction techniques in audio processing is the ability to apply SpecAugment techniques to the training spectrogram samples. In other words, frequency and time masking techniques were applied to the spectrograms in addition to time shifting, time stretching, and noise injection techniques that had been applied to the original training audio signals.

C. DATA NORMALIZATION TECHNIQUES

Another technique that can improve learning models is data scaling or normalization. The rescaling aims to normalize the values of spectrograms and extracted features so that they fall within a specific range. In this study, the normalization scaled the data values to the range of [0, 1]. By scaling, the model can learn more effectively and converge faster [47].

D. LEARNING ALGORITHMS

After the preprocessing techniques and preparing the audio signal to the suitable formats, the data can be fed into learning algorithms. Recently, CNN models have attracted scholar attention due to their ability to achieve high accuracy and extract features from data automatically. However, one of the complexities associated with deep learning models is their interpretability and explainability [48]. Deep learning models operate as black boxes, meaning it is challenging to understand the underlying patterns. On the other hand, machine learning models like the RF algorithms provide a transparent and interpretable framework for analyzing the features. RFs are well-suited for this study because they are less prone to overfitting due to imbalanced data, especially when the dataset size is not large enough to mitigate this risk [49]. Therefore, we employed two powerful learning approaches, namely CNNs and RFs to achieve both high accuracy and feature extraction explanations.

E. EVALUATION METRICS

To evaluate the performance of the fitted RFs and CNNs for the detection of cooling fan malfunctions, we employed the following key evaluation metrics [50]:

- Accuracy: it measures the overall correctness of the model predictions. It calculates the ratio of correctly classified samples to the total number of samples.
- Recall: it measures the ability of the models to correctly identify all instances of cooling fan malfunctions, providing insights into their capacity to minimize cases where malfunctioning cooling fans are not detected.
- F1 score: It balances the trade-off between the ability to correctly classify malfunctioning predictions and recall the ability to capture all malfunctioning instances.

F. TECHNIQUES TO IMPROVE LEARNING MODELS

This section outlines methods for improving the performance of learning models, including hyperparameter tuning, learning curve analysis, and classifier boundary threshold adjustment.

1) HYPERPARAMETER TUNING

Unlike the learning model parameters, which are set during training, hyperparameters are the parameters of learning algorithms that should be tuned before training. Since the performance of learning models highly depends on the hyperparameters of the model, tuning the hyperparameters is a crucial step in learning algorithms. In this study, Bayesian optimization, a probabilistic model-based optimization technique was employed to find the optimal set of hyperparameters. This systematic approach involves several vital steps [51]. First, we defined a search space encompassing the hyperparameters relevant to the learning models. Then, we selected accuracy as the objective function to quantify the performance of these models. Bayesian optimization began with initial random evaluations of this objective function to build an initial surrogate model. This surrogate model approximated the true objective function and guided the selection of hyperparameters to evaluate next, using an acquisition function that balanced exploration and exploitation. The process iteratively refined the surrogate model, optimizing the acquisition function until a predefined convergence criterion was met. Ultimately, Bayesian optimization provided us with the optimal set of hyperparameters, which we used to train the learning models.

2) LEARNING CURVE

A comprehensive learning curve analysis was conducted on the training and validation datasets to ensure the learning models were trained effectively on the dataset. It helps to identify underfitting, which occurs when the learning models are too simple to capture the underlying patterns within the data, and overfitting when the models are overly complex and essentially memorize the training data instead of generalizing from it. In addition, it provides insight into whether or not more training data should be acquired or regularization techniques should be employed to enhance the performance of the models.

To construct the learning curves for RFs, we trained the RFs initially with a single audio signal and subsequently evaluated them on the validation data. The training sample size was incrementally increased to cover the entire training dataset. For each training sample size, the learning curve was created by plotting the accuracy of both the training and validation data. However, in CNNs, the learning curve demonstrates how the loss of the model changes with respect to the number of epochs or iterations. An optimal learning curve is characterized by a small gap between the training and validation curves, indicating good generalization. Additionally, when both curves stabilize at a low loss (for CNNs) or high accuracy (for RFs), it suggests that learning algorithms perform well on the dataset. Such trends indicate that the models are generalizing effectively and adding further training data may not yield significant improvements. Conversely, if these characteristics are not observed in the learning curves, it is an indication that the model performance may be suboptimal and require further adjustment [52].

3) ADJUSTING THE CLASSIFICATION THRESHOLD

In addition to hyperparameters that considerably impact the learning classifier algorithms, the decision boundary for classifying samples must be optimized. The decision boundary for classifying samples into normal or malfunctioning classes in the learning algorithms is typically set at 0.5 by default. In other words, if the probability of a sample belonging to the normal class is greater than or equal to 0.5, it is classified as normal; otherwise, it will be classified as malfunctioning. However, the default threshold of 0.5 may not be optimal. The receiver operating characteristic (ROC) curve was used to determine the optimal threshold. ROC curves are graphical representations of the performance of a learning classifier across various thresholds. The Y-axis represents the true positive rate (TPR), and the X-axis represents the false positive rate (FPR). The TPR and FPR can be expressed as follows:

$$\text{TPR} = \frac{TP}{TP + FN} \tag{18}$$

$$\text{FPR} = \frac{FP}{FP + TN} \tag{19}$$

where TP and TN are the numbers of correctly classified normal and malfunctioning cases, whereas FP and FN are the counts of incorrectly classified malfunctioning and normal cases, respectively. The top-left corner of the ROC plot denotes the ideal point where the TPR is 1 and the FPR is 0. So, the optimal threshold is the point closest to the top-left corner of the plot [53].

G. FEATURE IMPORTANCE

RFs offer a transparent and interpretable framework for feature analysis. They create a hierarchical structure of decision rules that can be easily visualized and understood. In an RF model, nodes and branches represent specific features

and their corresponding thresholds, enabling a clear interpretation of the model decision-making process. To leverage RF capability to analyze features and enhance interpretability, we utilized the SHapley Additive exPlanations (SHAP) method [54]. SHAP values quantify the importance and contribution of the extracted time-frequency features to identify which acoustic features or characteristics of the cooling fan sounds are most influential in determining whether a malfunction is present.

H. COMPUTATIONAL COMPLEXITY

Computational complexity in the context of the learning models refers to the amount of computational resources needed for their training and operation.

1) CONVOLUTIONAL NEURAL NETWORK

The computational complexity during the training phase of a CNN primarily involves the forward pass and the backward pass. The complexity for each convolutional layer can be expressed as:

$$C^{T}_{CNN,Con} = D_{in} \times W_{in} \times H_{in} \times N \times D_{out} \times W_f \times H_f \tag{20}$$

where, $C^{T}_{CNN,Con}$ indicates the computational complexity of a convolutional layer during training, D_{in} is depth of input feature map, W_{in} and H_{in} are width and height of input feature map, N is number of filters in the layer, W_f and H_f are width and height of the filter or kernel, and D_{out} epresents depth of output feature map. For dense layers:

$$\mathrm{C^{T}_{CNN,Dense}} = \text{Input Neurons} \times \text{Output Neurons} \tag{21}$$

where, $C^{T}_{\mathrm{CNN,Dense}}$ indicates the computational complexity of a dense layer during training. Besides, during the prediction phase, only the forward pass is computed. Thus, the complexity is generally half of that in the training phase for each layer. It should be noted that dropout layers are typically inactive during the prediction phase, so their impact on complexity is generally only considered during training. Total computational complexity of the CNN is the summation of the operations from all layers.

2) RANDOM FOREST

RF complexity depends on the number of trees and the depth of each tree. The methodology for calculating computational complexity for training and prediction differs. Computational complexity during training could be calculated as follows [55]:

$$C^{T}_{RF} = M \times k \times n \times \log(n) \tag{22}$$

where, C^{T}_{RF} indicates the computational complexity of a RF during training, M is the number of trees, k *is*number of features, and n indicates the number of training sample. Computational complexity during prediction could be expressed as follows [55]:

$$C^{P}_{RF} = M \times k \tag{23}$$

where, C^{P}_{RF} indicates the computational complexity of a RF during prediction, M and k are the number of grown trees and the number of features, respectively. It should be noted that during the training phase of classification tasks with RF, when max_features is set to Auto, the algorithm typically uses the square root of the total number of features. This implies that in equation (22), square root of k is used instead of k.

Besides, the organizational flowchart of the simulation procedure in this study is shown in Figure 1.

III. RESULTS AND DISCUSSION

All the simulations were performed on a system equipped with an Intel(R) Core (TM) i7-8700 CPU @ 3.20GHz and 16 GB of memory.

A. DATA PREPROCESSING AND AUGMENTATION

This study employed time stretching, time shifting, and noise injection data augmentation techniques sequentially and parallelly to enrich the training dataset.

Time stretching was used to generate additional audio samples for both normal and malfunctioning operation of the fan sounds. By randomly adjusting the playback speed from 0.8 to 1.5, we generated new audio samples with lengths ranging from 3.4 to 6.4 seconds. The size of the training dataset was further increased by randomly shifting the original and stretched signals. Besides, artificially generated and original audio signals were rendered more realistic by exposing them to white and environmental noises. In this study, we assumed that 20 percent of the training data were randomly exposed to environmental and white noise. Figures 2 and 3 display the various data augmentation techniques used for a single instance of the normal and malfunctioning cooling fan operation. A stretch factor of 0.9 increased the original audio signal duration from 5.2 seconds to 5.8 seconds. This change created a slower auditory perception, simulating a cooling fan operating at a reduced speed while preserving the original pitch. Due to the use of small stretch factors, the shape of the signal was not distorted significantly, confirming the credibility of the time stretching method. Time shifting preserved amplitude, frequency content, and energy while displacing the signal by 0.5 seconds along the time axis to the right. Thundering noise has minimal impact on the original signal characteristics up to the 2-second. However, after this point, thundering noise significantly changed the amplitude, frequency components, and energy of the original signal, serving as a tangible example of how environmental factors can affect acoustic audio signals. White noise introduced fluctuations in amplitude, influenced the frequency content by spreading energy across all frequencies, and increased the signal's overall power, resulting in loudness variations. The combination of these data augmentation techniques significantly expanded the training dataset diversity.

B. FEATURE EXTRACTION AND SPECTROGRAM REPRESENTATION

Noticeable distinctions are observed in terms of amplitude, frequency, and spectral density in Figures 2 and 3. However,

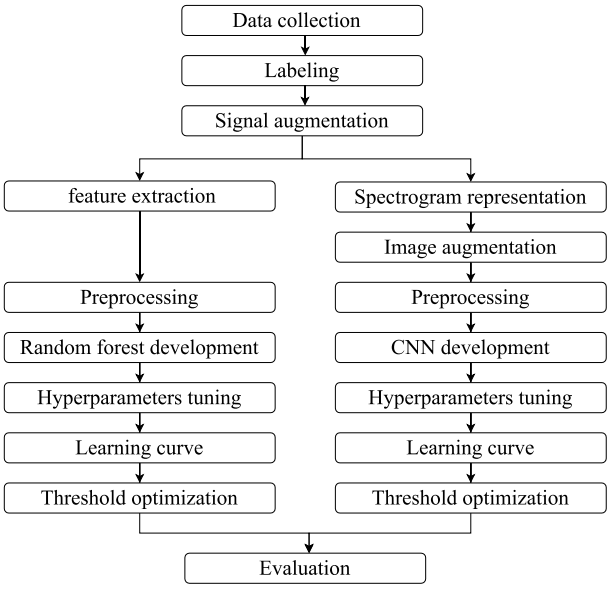


FIGURE 1. Organizational flowchart of the simulation procedure in this study.

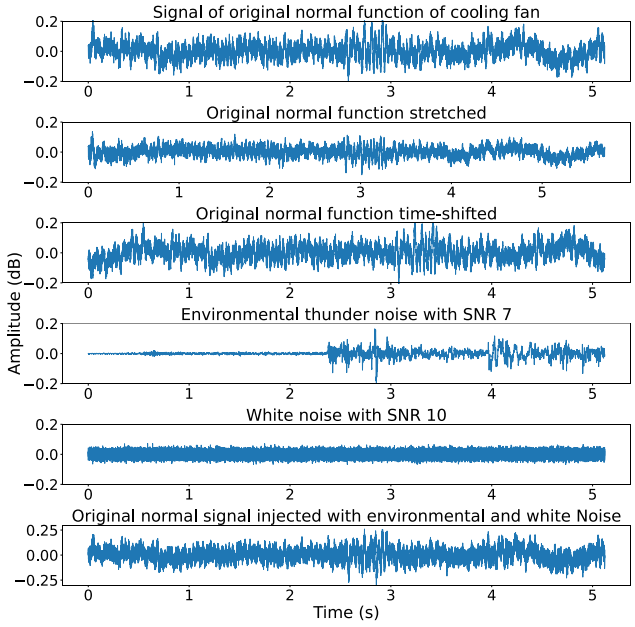


FIGURE 2. Various data augmentation techniques used for a single instance of the normal cooling fan audio signal.

we explored more specific audio characteristics in the following sections to better understand these differences for developing effective RF and CNN classifier models.

1) FEATURE EXTRACTION RESULTS

A comprehensive set of time and frequency features was extracted from normal and malfunctioning AF transformer audio signals to aid learning algorithms in discriminating between the two. For example, Figure 4 presents the extracted joint time-frequency features for the original normal and

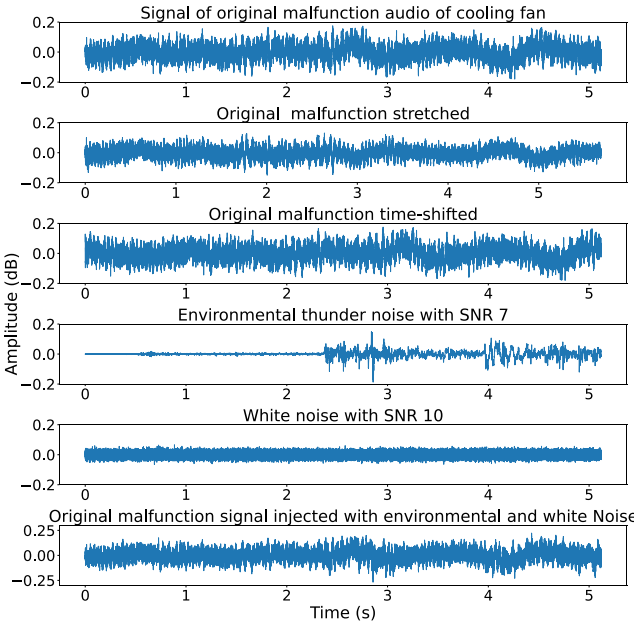


FIGURE 3. Various data augmentation techniques used for a single instance of the malfunctioning cooling fan audio signal.

malfunctioning audio signals depicted in Figures 2 and 3. Specific features such as ZCR, spectral centroid, and spectral bandwidth exhibit substantial differences between normal and malfunctioning audio signals. These differences indicate that normal and malfunctioning fans produce unique audio characteristics that offer valuable discriminative characteristics for effectively distinguishing between the two signals. In malfunctioning cooling fans, high ZCR values can be attributed to abrupt changes, transient events, or irregular sound patterns. However, the spectral centroid and bandwidth values are notably higher for the normal audio signals compared to the malfunctioning ones. The difference in spectral centroid can be attributed to the fact that normal cooling fans produce consistent and relatively steady sounds, resulting in a higher spectral centroid. The elevated spectral bandwidth in the normal signals indicates a broader range of frequencies in these audio samples. It is consistent with the typical behavior of healthy cooling fans, which produce sounds spanning various frequencies. In contrast, malfunctioning fans often generate audio signals with a narrower frequency range, leading to lower spectral bandwidth values. In addition, spectral flatness values are higher for normal audio, indicating that normal signals have a more uniform power distribution across the frequency spectrum. This aligns with the expected behavior of normal cooling fans, which produce audio signals with a balanced energy distribution across different frequencies. Further, the dominance of peak frequency in normal audio signals suggests that normal cooling fans tend to produce consistent frequency peaks in their signals. Notably, the variance for normal audio signals remains relatively steady with minor fluctuations, whereas malfunctioning audio signals exhibit significant volatility. Besides, the RMS and minimum values

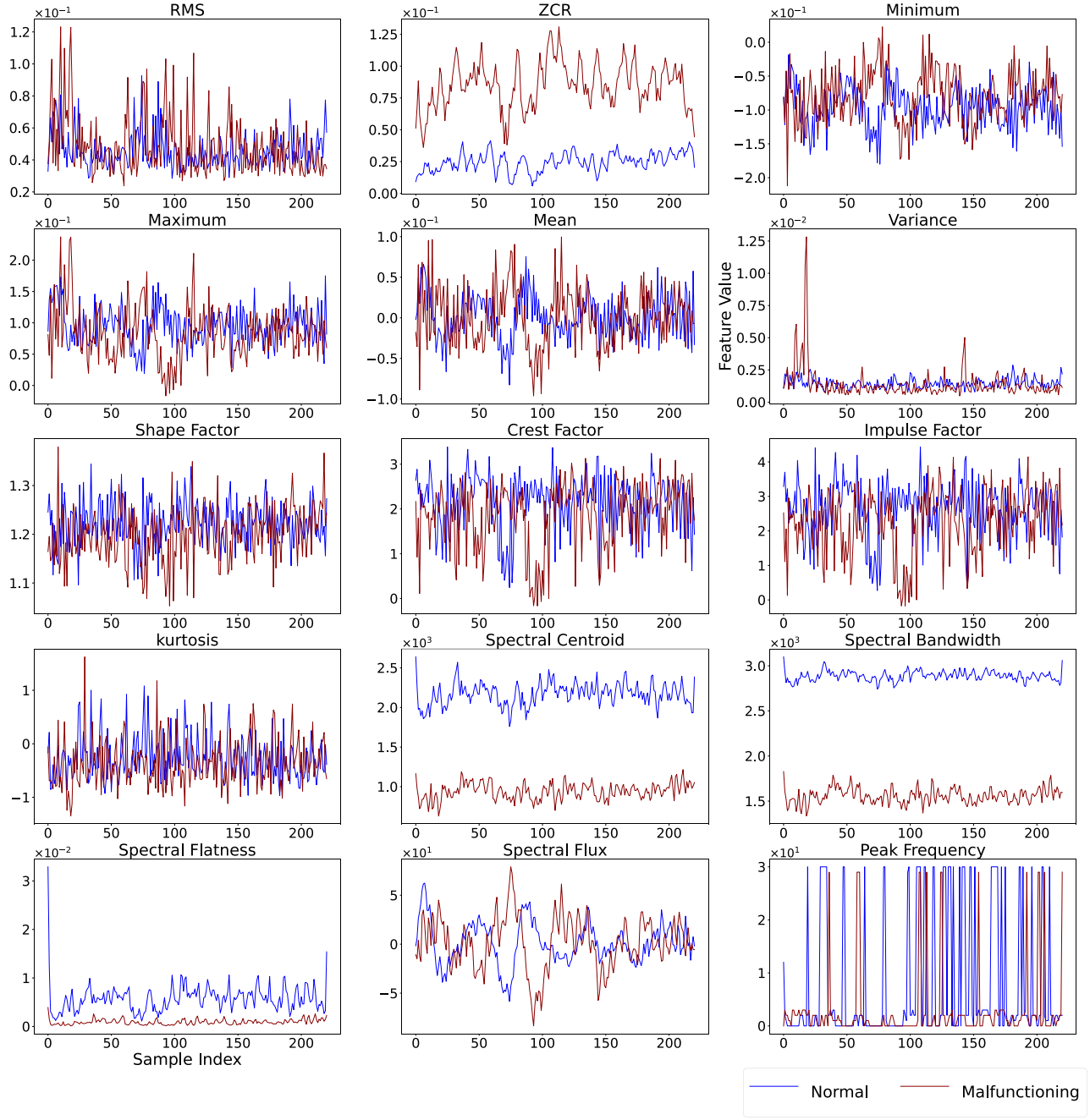


FIGURE 4. Derived joint time-frequency characteristics for the original normal and malfunctioning audio signals depicted in Figures 2 and 3.

are relatively higher for malfunctioning audio signals, which may indicate more pronounced sound intensity and possible irregularities in the malfunctioning audio signal. However, for the remaining features, no significant distinction can be observed between the audio signals of cooling fans. These distinctive extracted features from the audio signals of cooling fans were used as discriminators in classifying normal and malfunctioning AF transformer audio signals using the RFs.

2) SPECTROGRAM REPRESENTATION

We transformed the audio signals into spectrogram image representations to leverage the power of CNNs for AF transformer malfunctioning detection. Figure 5 displays original spectrograms alongside their time-frequency masked counterparts, derived from the original normal and malfunctioning audio signals shown in Figures 2 and 3. Darker shades indicate lower energy or amplitude, whereas brighter colors signify higher energy. The masked frequency and time

areas appear as horizontal and vertical bands of blackness, respectively, indicating zero energy or absence of information in those specific frequency and time regions. The spectrogram of the malfunctioning audio signal exhibits a brighter area than the normal one. This suggests a higher energy or amplitude in the malfunctioning signal, indicating potential irregularities or abnormalities in the frequency distribution. Besides, several bright frequency bands in low and high frequencies are observed in the malfunctioning audio signal, whereas only a few are seen for normal function. This indicates a significant concentration of energy at various frequencies in the normal and malfunctioning signal. These observations underscore the distinctions within the spectrograms, indicating their potential as effective discriminators for classifying normal and malfunctioning audio signals in AF transformer cooling fans using CNNs.

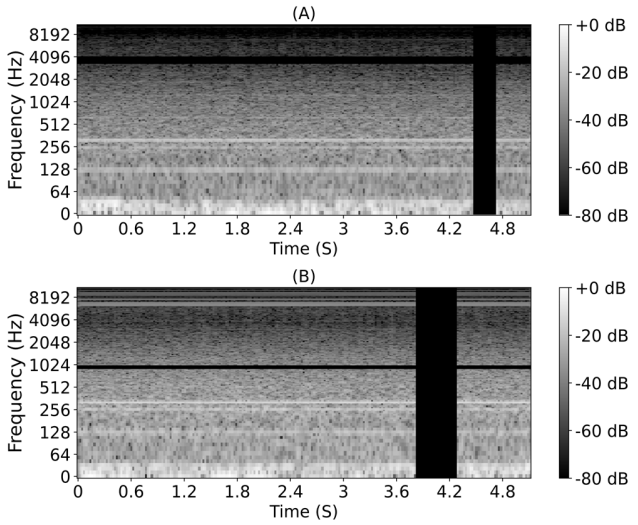


FIGURE 5. Spectrograms of augmented audio signals are depicted in Figures 2 and 3. (A) illustrates the frequency and time masked spectrogram of the original normal audio. (B) presents the frequency and time masked spectrogram of the original malfunctioning audio.

C. CNN CLASSIFICATION

A CNN classification algorithm was developed in this study to classify the spectrograms into normal and malfunctioning classes. The image pixels were normalized into values between 0 and 1 to achieve better convergence rates during training. Then, the CNN model was developed with multiple layers, including convolutional layers for feature extraction and max-pooling layers for dimensionality reduction.

Besides, batch normalization and dropout layers were implemented to improve model generalization ability and prevent overfitting. The optimal hyperparameters of the fitted CNN by using Bayesian optimization are given in Table 1. Additionally, the computational complexity of this CNN model is presented in Table 2. The computational complexity of the CNN model is approximately 31.69 million operations during training and 13.2 million operations in the prediction phase. This significant reduction in complexity during prediction is attributed to excluding specific processes active only in the training phase.

TABLE 1. Optimal hyperparameters of the fitted CNN.

Hyperparameters	Values
Number of convolutional layers	3
Number of dense layers	2 (fully connected)
Kernel size of the first layers	3×3
Number of neurons in the first dense layer	256
Number of neurons in the second dense layer	8
Activation function of the layers	Relue and sigmoid
Pooling size after the first layer	2×2
Pooling size after the second layer	2×2
Number of filters in first convolutional layer	64
Number of filters in second convolutional layer	128
Number of filters in third convolutional layer	256
Dropout rate	0.5
Batch size	32
Loss function	Categorical cross-entropy
Learning rate	0.001
Learning rule	Adam

TABLE 2. Computational complexity analysis of the fitted CNN and random forests.

Model	Training complexity (operations)	Prediction complexity (operations)
CNN	31.69 Million	13.2 Million
RF	7.9 Million	0.66 Million

1) CNN LEARNING CURVE ANALYSIS

Figure 6 demonstrates the learning curve of the trained CNN by using the loss of the training and validation datasets. Initial training and validation losses are relatively high. This indicates that the model started with a high error rate, which is expected as the model was initialized with random weights. The training losses decreased as the epochs increased and converged to a stable point, indicating that the CNN learned to fit the training datasets. A similar trend in the validation set suggests that the model generalized well to validation data but shows some fluctuations. These fluctuations may be due to the model facing new patterns in the validation set that it had not encountered during training. The training and validation curves converged after epochs around 20 and beyond this point, with both losses fluctuating slightly but staying relatively constant. Accordingly, further adding more training data is unlikely to yield significant improvements. In addition, the gap between the training and validation loss is shallow, representing that the fitted CNN does not suffer from overfitting or underfitting. Consequently, the fitted CNN performs well and can capture the underlying patterns of the audio signals.

2) CNN OPTIMAL CLASSIFIER THRESHOLD ANALYSIS

Besides, the CNN model was executed multiple times with various thresholds, and in each iteration, the FPR and TPR were calculated. The resulting ROC curve is depicted in

Figure 7. The diagonal line serves as a baseline and denotes a random classifier with no discriminative ability between the normal and malfunctioning classes. ROC curves represent CNN performance at various thresholds. This curve above the diagonal line indicates the ability of CNNs to distinguish between the normal and malfunctioning classes. The closest point to the top-left corner of the ROC plot occurs at a TPR of 0.96 and an FPR of 0.06. This point suggests that the model correctly identifies 96% of the actual normal audio signals and only 6% of the malfunctioning signals are incorrectly classified as normal, which is relatively low. The high TPR of 0.96 means the CNNs have a superior ability to detect the normal class. The low FPR of 0.06 implies the model has low false alarms and rarely misclassifies malfunctioning instances as normal. This is crucial in operational settings to avoid unnecessary maintenance or inspections. This optimal point corresponds to a classification threshold of 0.44. That means CNNs classify instances as normal if the probability of them being normal is greater than or equal to 0.44; Otherwise, they are classified as malfunctioning.

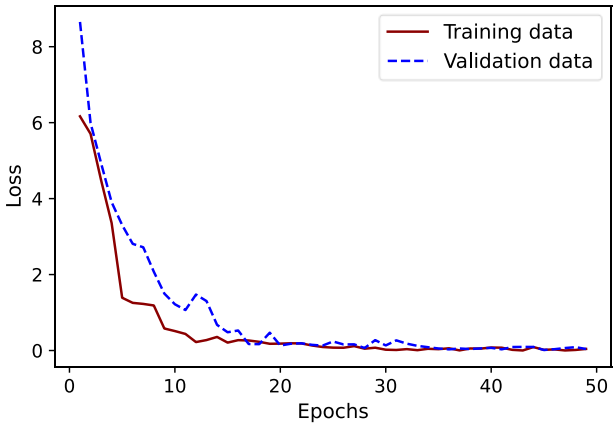


FIGURE 6. Learning curve associated with the convolutional neural network.

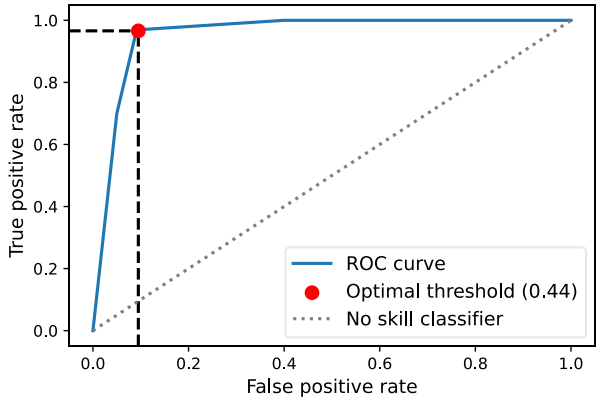


FIGURE 7. Receiver operating characteristic curve associated with CNN classifier.

To evaluate the impact of the optimal threshold, Table 3 presents the performance metrics of the CNN model using both the optimal threshold of 0.44 and the default threshold of 0.5. The optimal threshold of 0.44 results in superior performance to the commonly used threshold of 0.5 across all metrics. These results underscore the importance of optimizing the classification threshold when developing the learning models.

TABLE 3. CNN performance with different threshold.

Threshold	Metric	Training	Test
0.44	Accuracy	0.94	0.97
	Recall	0.96	0.96
	F1 score	0.95	0.98
0.5	Accuracy	0.89	0.92
	Recall	0.9	0.89
	F1 score	0.92	0.94

Based on the evaluation of the CNN model using the optimal threshold of 0.44, high values of F1 score, accuracy, and recall for both the test and training datasets exhibit the robust performance of the CNN to classify the cooling fan normal and malfunction operation. Model accuracy on unseen data is remarkably high at 97%, indicating that CNN can classify 97% of the audio signals in the test dataset correctly. A recall of 0.96 on the test set means that the CNN model correctly identified 96% of the actual normal cooling fan instances. Besides, an F1 Score of 0.98 indicates that the model effectively identified normal and malfunctioning cooling fan states, with very few classification errors. As a result of the fitted CNN on the entire training sample size, the model can be trusted where accurate classification of cooling fan operational states (evidenced by the high accuracy), identifying normal fans (evidenced by the high recall), and avoiding false alarms (high F1 score) are critical. Furthermore, the slight differences in performance metrics between the training and test datasets could be attributed to the extensive data augmentation applied to the training datasets. That is, the test datasets which are not subject to the same level of variability, the model could classify the data more efficiently.

3) DATA AUGMENTATION ANALYSIS

In addition, to precisely assess the impact of data augmentation techniques on model performance and their ability to capture hidden patterns in audio signals, CNNs were trained with the optimal threshold on the non-augmented data, 50%, 75%, and 100% of the entire training samples. Results using 100% of the data are those in Table 3, and for non-augmented data, 50%, and 75% of the training samples were provided in Table 4. Model performance drops as the training sample size decreases. When only the non-augmented training data was used, performance metrics decreased noticeably. Due to the limited diversity and size of the raw dataset, the model may not be effectively trained to generalize well to unseen data. However, model performance improves considerably as the training sample size increases to 50%, 75%, and 100%, all through data augmentation techniques. This improvement

suggests that data augmentation plays a significant role in enhancing model performance and generalization ability.

The consistently high performance of the model trained on 100% of the current augmented training dataset indicates that the existing level of data augmentation is acceptable. The model has achieved substantial results across multiple evaluation metrics. This suggests that the model has effectively learned the complexities of the audio signals and can be utilized for malfunction detection in AF transformers.

TABLE 4. CNN performance on various training sample sizes.

Training sample size	Metric	Training
75%	Accuracy	0.90
	Recall	0.88
	F1 score	0.92
50%	Accuracy	0.77
	Recall	0.73
	F1 score	0.8
Non-augmented	Accuracy	0.58
	Recall	0.48
	F1 score	0.63

TABLE 5. Parametric values of the random forest.

Hyperparameters	Values
n_estimators	200
max_depth	50
min_samples_split	20
min_samples_leaf	1
max_features	Auto
min_impurity_decrease	0.0
bootstrap	True
class_weight	2×2
ccp_alpha	0.0
criterion	entropy
warm_start	True
Learning rate	0.001
Learning rule	Adam

D. RANDOM FOREST CLASSIFICATION

In this section, the RF classifier was employed to classify the audio signals of the AF transformer. Like CNN, all preprocessing techniques (except SpecAugment techniques) were used to prepare the samples for further analysis. To train RFs, features were extracted from a combination of time and frequency-domain methods. The Bayesian optimization was applied to find the optimal hyperparameters of the RF algorithm. Table 5 provides the architecture and optimal hyperparameters of the RFs by utilizing Bayesian optimization. Additionally, Table 2 gives computational complexity to the RF model. In the prediction phase for the RF model, the computational complexity is notably lower. This is because the model primarily traverses existing trees for classification purposes rather than constructing new trees. Consequently, the computational demand in the prediction phase is markedly lower than in the training phase.

1) RF LEARNING CURVE ANALYSIS

The learning curve of the RF classifier is shown in Figure 8. The RF model exhibits perfect accuracy with a single training sample. This is expected given that the RFs have been trained on a singular data point and can predict it accurately. In contrast, validation accuracy is significantly low at this stage. When trained on multiple samples, the model fails to generalize to the diverse patterns in the validation dataset. As the training sample size increases, the RF model encounters more varied data, making perfect accuracy harder to achieve in the training dataset. This slight reduction in training accuracy signifies the model shift from memorizing data to generalizing across varied patterns. In addition, as the training sample size increases, the validation accuracy rises, indicating that RF models can generalize and predict unseen data more accurately. However, after 6-7 steps, both curves become relatively stable, suggesting that further data may not yield significant improvements. Furthermore, the small gap between the training and validation curves indicates that the model is complex enough to capture the underlying patterns in the data. So, achieving high accuracy without overfitting or underfitting suggests that the developed RF classifier serves as a highly reliable and effective model for the malfunctioning detection of AF transformer cooling fans.

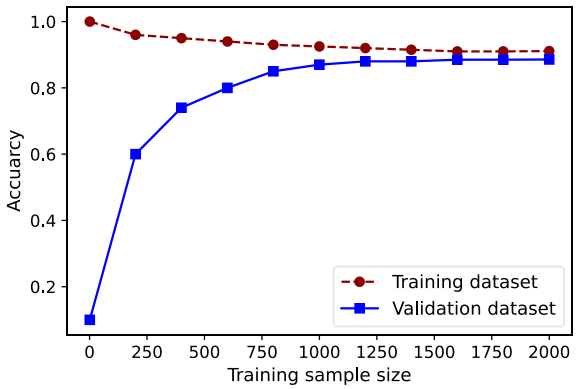


FIGURE 8. Learning curve of the random forests.

2) RF OPTIMAL CLASSIFIER THRESHOLD ANALYSIS

The next essential step is identifying the optimal threshold for RF classifiers to improve performance. Similar to the methodology employed for the CNN, the corresponding FPR and TPR were calculated and presented in Figure 9 for each threshold. The optimal threshold for the RF classifier is 0.39, corresponding to a TPR of 0.91 and an FPR of 0.12. With a high TPR of 0.91, the model demonstrates strong capability in correctly identifying normal cooling fans. Simultaneously, the relatively low FPR of 0.12 indicates a minimized false alarm rate. The performance of the RF classifier under both optimal and default thresholds is summarized in Table 6. The comparison confirms that by employing the optimal threshold of 0.39, the model yields accuracy, recall, and F1 scores better than the default threshold of 0.5. Therefore, the threshold

of 0.39 makes the RF classifier more effective for detecting malfunctioning cooling fans than the default threshold of 0.5. These high levels of accuracy, F1 score, and recall indicate that the model effectively classifies the cooling fan operational status as normal or malfunctioning.

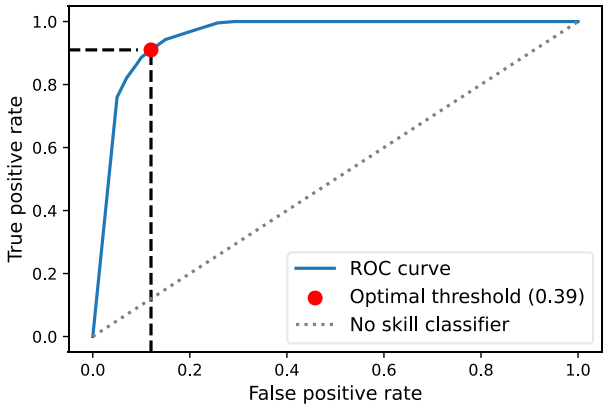


FIGURE 9. Receiver operating characteristic curve associated with random forest classifier.

TABLE 6. Random forest performance with different thresholds.

Threshold	Metric	Training	Test
0.39	Accuracy	0.93	0.91
	Recall	0.93	0.91
	F1 score	0.93	0.92
0.5	Accuracy	0.89	0.92
	Recall	0.9	0.89
	F1 score	0.92	0.94

3) FEATURE IMPORTANCE ANALYSIS

Although the RF classifier performs lower than CNNs, it offers a special advantage in feature interpretability and explainability that is rarely achievable with CNNs. Figure 10 demonstrates this capability by presenting the SHAP values, in which each feature importance was measured by its mean absolute value across all samples. Among the various time-frequency features utilized for classifying AF transformer cooling fan operating states, spectral bandwidth stands as the most significant feature. Its high importance in both classes reveals that the spread of frequencies in audio signals is a crucial discriminant in classifying fan operational status. Following the spectral bandwidth, the second important feature is the spectral centroid, indicating that malfunctioning fan and healthy fan audio signals have a distinct center of mass in the frequency distribution. For instance, a malfunctioning fan may produce a different whining or buzzing sound than a healthy fan, resulting in a different distribution of frequencies in the audio signals. Spectral flatness, peak frequency, and ZCR are the following significant features, with relatively the same importance values. The slightly higher SHAP value of peak frequency for normal fans compared to malfunctioning fans indicates that the dominant frequency in the audio signals of normal fans is slightly more influential in classification than the audio signals of malfunctioning fans. A high SHAP value of ZCR for malfunctioning fans compared to normal fans implies that malfunctioning cooling fans in AF transformers produce signals with more abrupt amplitude changes. However, the remaining features contribute less to the classification task. RMS, minimum, variance, Kurtosis factor, spectral flux, shape factor, and maximum have moderate or insignificant contributions. The crest factor, impulse factor, and mean have the least contribution. This means that the specific shape of the waveform, impulsiveness, and average amplitude level of the signals are not solid indicators for classifying fan statuses.

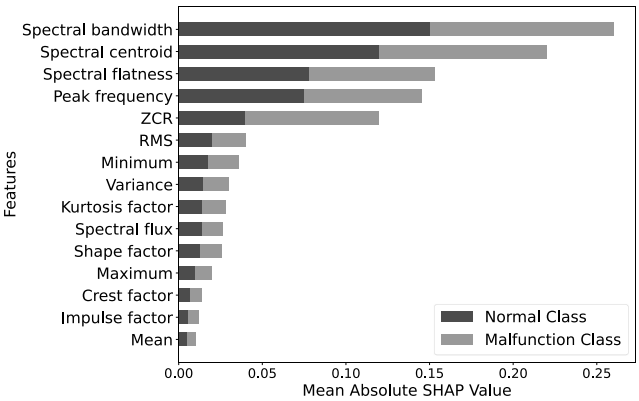


FIGURE 10. Mean absolute SHAP values of the extracted features.

In the task of malfunction detection of AF transformer cooling fans, the CNN outperforms the RF classifier across accuracy, recall, and F1 score. This superior performance can be attributed to the following factors. First, CNNs can learn hierarchical features automatically from raw data, whereas RFs are limited by manually extracted features, potentially missing the characteristic and relevant features in the data. Secondly, the deep architecture of CNNs is capable of capturing complex, non-linear relationships, a capability that may be limited in RF. Third, CNNs benefit from a wider range of data augmentation techniques than RFs by time and frequency masking techniques on the spectrograms. This enhances CNNs learning from a more diverse dataset.

E. TESTING THE CNN MODEL ON A NEW AF-TRANSFORMER

To test the fitted CNN model with an optimal classifier threshold of 0.44 in real-world scenarios, it was applied to a new dataset comprising audio recordings from another AF transformer situated in a non-isolated room environment in an urban area. This setting introduces a diverse range of ambient noises and operational variations, providing a more rigorous testing ground for the model applicability. The dataset included 20 normal and 15 malfunctioning recordings, offering a balanced representation of operational states. The results of this test, presented in Table 7, reveal the fitted CNN model sustained high performance in a varied

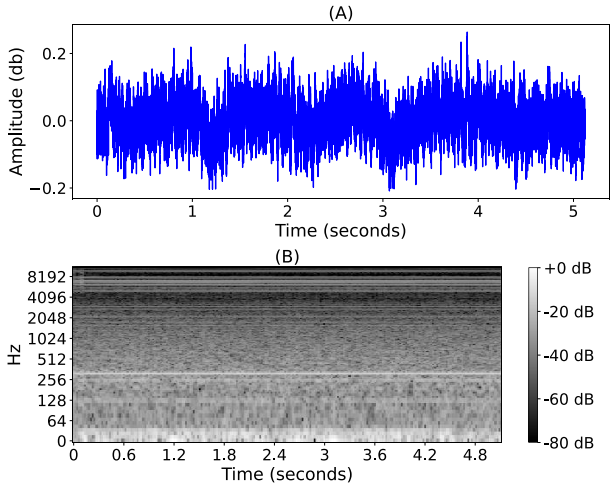


FIGURE 11. Analysis of misclassified normal audio. (A) represents a sample of normal audio that was incorrectly classified as malfunctioning. (B) shows the corresponding spectrogram of the same normal audio sample.

TABLE 7. Performance metrics of the CNN model on audio recordings from a new af transformer.

Metric	Value
Accuracy	0.97
Recall	0.95
F1 score	0.975
True positive rate (TPR)	0.95
False positive rate (FPR)	0.0

dataset. The model achieved an accuracy of 0.97, which means it correctly identified 97% of the audio recordings as either normal or malfunctioning. Furthermore, an F1 Score of 0.975 suggests the model is reliable in its classifications and makes few errors. Remarkably, the FPR of zero indicates the exceptional accuracy of the model in detecting all malfunctioning operations of the AF transformer. This is crucial in predictive maintenance, ensuring no potential issues are overlooked, thus preventing equipment failure and minimizing downtime. The recall, also known as the TPR, is 0.95. This high recall rate demonstrates the efficiency of the model in correctly identifying 95% of the actual malfunctioning recordings, with only one normal functioning recording classified as malfunctioning. This normal audio and its corresponding spectrogram that was misclassified as malfunctioning is shown in Figure 11. The misclassification of the normal audio can be attributed to its resemblance to malfunctioning audio characteristics, as seen in Figure 3. This signal appears to be noisier, possibly due to new ambient noise in the testing environment, leading to confusion in the CNN model. Additionally, the spectrogram of this signal shows horizontal light bands similar to those observed in malfunctioning audios, indicating higher energy in specific frequency ranges, which are typical features of malfunctioning signals. Despite this, the overall performance of the fitted CNNs in a new challenging dataset underscores its effectiveness in real-world applications. It demonstrates the potential of the fitted model in accurately identifying the operational states of AF transformer cooling fans and significantly reducing the likelihood of false alarms.

IV. CONCLUSION

This study introduced an acoustic-based online monitoring approach for detecting malfunctions in cooling fans of AF dry-type transformers, addressing a particular gap in current monitoring systems. We collected audio data from an AF transformer cooling fan using a single microphone and employed CNNs and RF classifiers for malfunctioning detection. The limitations of the collected dataset were mitigated by applying various data augmentation techniques. Time-frequency domain feature extraction methods provided input to the RF models, whereas the CNNs exploited spectrograms. Moreover, the performance of these learning algorithms was enhanced through hyperparameter tuning and classification boundary threshold optimization using ROC methods. The simulation results revealed that the CNN classifier yielded a high accuracy of 97% and a high recall and f1 score of 0.96 and 0.98, respectively, indicating a substantiating ability for early malfunction detection. In addition, testing the fitted CNN model on a new dataset in a real-world environment further validated its reliability in practical applications. Although the RF classifier presented marginally lower performance metrics, its capability in feature interpretability provided insight into the extracted features. The findings of this study suggest several areas for future research:

- Considering the important features extracted in this study, explore different feature extraction techniques to improve RF performance.
- Investigating ensemble learning methods, including combining CNNs and RF algorithms.
- Analyzing the effect of varying audio signal lengths on detecting malfunctions.
- Evaluation of different data augmentation strategies, such as ensemble methods.
- Incorporating multi-modal sensor data, including vibration and temperature sensors, to develop a more comprehensive system.

REFERENCES

[1] A. R. Abbasi, "Fault detection and diagnosis in power transformers: A comprehensive review and classification of publications and methods," *Electric Power Syst. Res.*, vol. 209, Aug. 2022, Art. no. 107990.
[2] C. Hackl, J. Kullick, and N. Monzen, "Generic loss minimization for nonlinear synchronous machines by analytical computation of optimal reference currents considering copper and iron losses," in *Proc. 22nd IEEE Int. Conf. Ind. Technol. (ICIT)*, vol. 1, Mar. 2021, pp. 1348–1355.
[3] M. Aslam, I. U. Haq, M. S. Rehan, F. Ali, A. Basit, M. I. Khan, and M. N. Arbab, "Health analysis of transformer winding insulation through thermal monitoring and fast Fourier transform (FFT) power spectrum," *IEEE Access*, vol. 9, pp. 114207–114217, 2021.
[4] C. Lei, S. Bu, Q. Wang, N. Zhou, L. Yang, and X. Xiong, "Load transfer optimization considering hot-spot and top-oil temperature limits of transformers," *IEEE Trans. Power Del.*, vol. 37, no. 3, pp. 2194–2208, Jun. 2022.
[5] H. Amiri, "Analysis and comparison of actual behavior of oil-type and dry-type transformers during lightning," in *Proc. 25th Electr. Power Distrib. Conf. (EPDC)*, Aug. 2021, pp. 1–4.

[6] M. S. Mahdi, "Effect of fin geometry on natural convection heat transfer in electrical distribution transformer: Numerical study and experimental validation," *Thermal Sci. Eng. Prog.*, vol. 14, Dec. 2019, Art. no. 100414.

[7] M. Ngo, Y. Cao, D. Dong, R. Burgos, K. Nguyen, and A. Ismail, "Forced air-cooling thermal design methodology for high-density, high-frequency, and high-power planar transformers in 1U applications," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 11, no. 2, pp. 2015–2028, Apr. 2023.

[8] D. Wang, "A new testing method for the dielectric response of oil-immersed transformer," *IEEE Trans. Ind. Electron.*, vol. 67, no. 12, pp. 10833–10843, Dec. 2020.

[9] S. Zhao, Q. Liu, M. Wilkinson, G. Wilson, and Z. Wang, "A reduced radiator model for simplification of ONAN transformer CFD simulation," *IEEE Trans. Power Del.*, vol. 37, no. 5, pp. 4007–4018, Oct. 2022.

[10] Y. Biçen and F. Aras, "Smart asset management system for power transformers coupled with online and offline monitoring technologies," *Eng. Failure Anal.*, vol. 154, Dec. 2023, Art. no. 107674.

[11] L. Wang, W. Zuo, Z.-X. Yang, J. Zhang, and Z. Cai, "A method for fans' potential malfunction detection of ONAF transformer using top-oil temperature monitoring," *IEEE Access*, vol. 9, pp. 129881–129889, 2021.

[12] M. J. Zideh, P. Chatterjee, and A. K. Srivastava, "Physics-informed machine learning for data anomaly detection, classification, localization, and mitigation: A review, challenges, and path forward," *IEEE Access*, vol. 12, pp. 4597–4617, 2024.

[13] J. Parvez, "Real-time monitoring system of power transformer using IoT and GSM," *J. Mech. Continua Math. Sci.*, vol. 16, no. 9, pp. 1–15, Sep. 2021.

[14] V. Shiravand, J. Faiz, M. H. Samimi, and M. Mehrabi-Kermani, "Prediction of transformer fault in cooling system using combining advanced thermal model and thermography," *IET Gener., Transmiss. Distrib.*, vol. 15, no. 13, pp. 1972–1983, Jul. 2021.

[15] M. Zouiti, O. Bonnard, R. Desquiens, M. Cuevas, and D. Bortolotti, "Online monitoring of power transformers to improve their operating and maintenance model," in *Proc. CIRED 26th Int. Conf. Exhib. Electr. Distrib.*, vol. 2021, Sep. 2021, pp. 496–499.

[16] O. Laayati, M. Bouzi, and A. Chebak, "Design of an oil immersed power transformer monitoring and self diagnostic system integrated in smart energy management system," in *Proc. 3rd Global Power, Energy Commun. Conf. (GPECOM)*, Oct. 2021, pp. 240–245.

[17] H. Zhang, G. Liu, B. Lin, H. Deng, Y. Li, and P. Wang, "Thermal evaluation optimization analysis for non-rated load oil-natural air-natural transformer with auxiliary cooling equipment," *IET Gener., Transmiss. Distrib.*, vol. 16, no. 15, pp. 3080–3091, Aug. 2022.

[18] M. Djamali and S. Tenbohlen, "A validated online algorithm for detection of fan failures in oil-immersed power transformers," *Int. J. Thermal Sci.*, vol. 116, pp. 224–233, Jun. 2017.

[19] J. Picaut, A. Can, N. Fortin, J. Ardouin, and M. Lagrange, "Low-cost sensors for urban noise monitoring networks—A literature review," *Sensors*, vol. 20, no. 8, p. 2256, Apr. 2020.

[20] P. Thanapol, K. Lavangnananda, P. Bouvry, F. Pinel, and F. Leprévost, "Reducing overfitting and improving generalization in training convolutional neural network (CNN) under limited sample sizes in image recognition," in *Proc. 5th Int. Conf. Inf. Technol. (InCIT)*, Oct. 2020, pp. 300–305.

[21] V. Ravi, J. Wang, J. Flint, and A. Alwan, "Fraug: A frame rate based data augmentation method for depression detection from speech signals," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 6267–6271.

[22] P. Byun and J.-H. Chang, "Effective masking shapes based robust data augmentation for acoustic scene classification," in *Proc. 8th IEEE Int. Conf. Netw. Intell. Digit. Content (IC-NIDC)*, Nov. 2023, pp. 404–408.

[23] L. K. Shahidi, L. M. Collins, and B. O. Mainsah, "Parameter tuning of time-frequency masking algorithms for reverberant artifact removal within the cochlear implant stimulus," *Cochlear Implants Int.*, vol. 23, no. 6, pp. 309–316, Nov. 2022.

[24] A. Abeysinghe, S. Tohmuang, J. L. Davy, and M. Fard, "Data augmentation on convolutional neural networks to classify mechanical noise," *Appl. Acoust.*, vol. 203, Feb. 2023, Art. no. 109209.

[25] Q. Wang, J. Du, H.-X. Wu, J. Pan, F. Ma, and C.-H. Lee, "A four-stage data augmentation approach to ResNet-conformer based acoustic modeling for sound event localization and detection," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 31, pp. 1251–1264, 2023.

[26] M. Goubeaud, P. Joußen, N. Gmyrek, F. Ghorban, and A. Kummert, "White noise windows: Data augmentation for time series," in *Proc. 7th Int. Conf. Optim. Appl. (ICOA)*, May 2021, pp. 1–5.

[27] J. Chen, W. Yi, and D. Wang, "Filter bank sinc-ShallowNet with EMD-based mixed noise adding data augmentation for motor imagery classification," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Nov. 2021, pp. 5837–5841.

[28] N. Yalta, S. Watanabe, T. Hori, K. Nakadai, and T. Ogata, "CNN-based multichannel end-to-end speech recognition for everyday home environments," in *Proc. 27th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2019, pp. 1–5.

[29] C.-I. Lai, Y.-S. Chuang, H.-Y. Lee, S.-W. Li, and J. Glass, "Semi-supervised spoken language understanding via self-supervised speech and language model pretraining," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 7468–7472.

[30] Y. Cao, Y. Ji, Y. Sun, and S. Su, "The fault diagnosis of a switch machine based on deep random forest fusion," *IEEE Intell. Transp. Syst. Mag.*, vol. 15, no. 1, pp. 437–452, Jan. 2023.

[31] K. M. M. Prabhu, *Window Functions and Their Applications in Signal Processing*. Boca Raton, FL, USA: CRC Press, 2018.

[32] J. Chen, B. Xu, and X. Zhang, "A vibration feature extraction method based on time-domain dimensional parameters and Mahalanobis distance," *Math. Problems Eng.*, vol. 2021, pp. 1–12, Jul. 2021.

[33] G. Sharma, K. Umapathy, and S. Krishnan, "Trends in audio signal feature extraction methods," *Appl. Acoust.*, vol. 158, Jan. 2020, Art. no. 107020.

[34] S. Kavitha and J. Manikandan, "Improved methodology of SVM to classify acoustic signal by spectral centroid," *J. Trends Comput. Sci. Smart Technol.*, vol. 3, no. 4, pp. 294–304, May 2022.

[35] M. Lagrange and F. Gontier, "Bandwidth extension of musical audio signals with no side information using dilated convolutional neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 801–805.

[36] D. Aiordachioaie, "On feature selection from time-frequency images," in *Proc. 14th Int. Conf. Electron., Comput. Artif. Intell. (ECAI)*, Jun. 2022, pp. 1–4.

[37] C. Zhao, H. Wang, H. Chen, W. Shi, and Y. Feng, "JAMSNet: A remote pulse extraction network based on joint attention and multi-scale fusion," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 6, pp. 2783–2797, Jun. 2023.

[38] X. Zhang, Z. Su, P. Lin, Q. He, and J. Yang, "An audio feature extraction scheme based on spectral decomposition," in *Proc. Int. Conf. Audio, Lang. Image Process.*, Jul. 2014, pp. 730–733.

[39] Y. Liu, Z. Yan, A. Wu, T. Ye, and Y. Li, "Nighttime image dehazing based on variational decomposition model," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 639–648.

[40] Y. Liu, Z. Yan, S. Chen, T. Ye, W. Ren, and E. Chen, "NightHazeFormer: Single nighttime haze removal using prior query transformer," in *Proc. 31st ACM Int. Conf. Multimedia*. New York, NY, USA: ACM, Oct. 2023, pp. 4119–4128.

[41] Y. Liu, Z. Yan, J. Tan, and Y. Li, "Multi-purpose oriented single nighttime image haze removal based on unified variational retinex model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 4, pp. 1643–1657, Apr. 2023.

[42] Y. Jin, B. Lin, W. Yan, Y. Yuan, W. Ye, and R. T. Tan, "Enhancing visibility in nighttime haze images using guided APSF and gradient adaptive convolution," in *Proc. 31st Int. Conf. Multimedia*, New York, NY, USA: ACM, Oct. 2023, pp. 2446–2457.

[43] Y. Liu, Z. Yan, T. Ye, A. Wu, and Y. Li, "Single nighttime image dehazing based on unified variational decomposition model and multi-scale contrast enhancement," *Eng. Appl. Artif. Intell.*, vol. 116, Nov. 2022, Art. no. 105373.

[44] Z. Sun, Q. Ke, H. Rahmani, M. Bennamoun, G. Wang, and J. Liu, "Human action recognition from various data modalities: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 3, pp. 3200–3225, Mar. 2023.

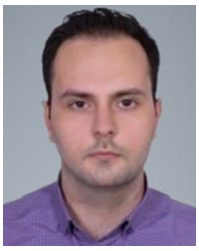
[45] D. S. Park, W. Chan, Y. Zhang, C.-C. Chiu, B. Zoph, E. D. Cubuk, and Q. V. Le, "SpecAugment: A simple data augmentation method for automatic speech recognition," in *Proc. Interspeech*, Sep. 2019, pp. 2613–2617.

[46] Y. Ren, D. Liu, C. Liu, Q. Xiong, J. Fu, and L. Wang, "A universal audio steganalysis scheme based on multiscale spectrograms and DeepResNet," *IEEE Trans. Depend. Secure Comput.*, vol. 20, no. 1, pp. 665–679, Jan. 2023.

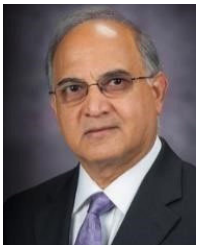
[47] M. A. Siddiqi and W. Pak, "An agile approach to identify single and hybrid normalization for enhancing machine learning-based network intrusion detection," *IEEE Access*, vol. 9, pp. 137494–137513, 2021.

[48] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable AI: A review of machine learning interpretability methods," *Entropy*, vol. 23, no. 1, p. 18, 2020.
[49] M. P. Neto and F. V. Paulovich, "Explainable matrix–visualization for global and local interpretability of random forest classification ensembles," *IEEE Trans. Vis. Comput. Graphics*, vol. 27, no. 2, pp. 1427–1437, Feb. 2021.
[50] A. Yousaf, M. Umer, S. Sadiq, S. Ullah, S. Mirjalili, V. Rupapara, and M. Nappi, "Emotion recognition by textual tweets classification using voting classifier (LR-SGD)," *IEEE Access*, vol. 9, pp. 6286–6295, 2021.
[51] H. Cho, Y. Kim, E. Lee, D. Choi, Y. Lee, and W. Rhee, "Basic enhancement strategies when using Bayesian optimization for hyperparameter tuning of deep neural networks," *IEEE Access*, vol. 8, pp. 52588–52608, 2020.
[52] T. Viering and M. Loog, "The shape of learning curves: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 6, pp. 7799–7819, Jun. 2023.
[53] D.-K. To, G. Adimari, M. Chiogna, and D. Risso, "Receiver operating characteristic estimation and threshold selection criteria in three-class classification problems for clustered data," *Stat. Methods Med. Res.*, vol. 31, no. 7, pp. 1325–1341, Jul. 2022.
[54] T.-T.-H. Le, H. Kim, H. Kang, and H. Kim, "Classification and explanation for intrusion detection system based on ensemble trees and SHAP method," *Sensors*, vol. 22, no. 3, p. 1154, 2022.
[55] K. Hassine, A. Erbad, and R. Hamila, "Important complexity reduction of random forest in multi-classification problem," in *Proc. 15th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2019, pp. 226–231.

REZA NEMATIRAD (Graduate Student Member, IEEE) received the B.Sc. degree in electrical engineering and the M.Sc. degree in power system planning and management from the Center of Excellence in Power Systems, Amirkabir University of Technology, Tehran, Iran, in 2018 and 2021, respectively. He is currently pursuing the Ph.D. degree in electrical and computer engineering with Kansas State University. His research interests include optimization in power systems, state estimation, probabilistic and statistical analysis, renewable energy resources, learning algorithms, and time series analysis.



MEHDI BEHRANG received the B.S. degree in electrical engineering from the Amirkabir University of technology, Tehran, Iran, in 2020, and the M.S. degree in electrical and electronic engineering from Sheffield Hallam University, Sheffield, U.K., in 2022. He is currently pursuing the Ph.D. degree in electrical and computer engineering with Carleton University, Ottawa, ON, Canada. From 2020 to 2022, he was a Research Assistant with the Artificial Intelligence Laboratory, Sheffield Hallam University. His research interest includes the development of novel loss functions for one-stage object detector algorithms, such as RetinaNet, and the improvement of computer vision algorithms. In his research, he tried to solve the imbalance class problem using focal loss function in different architectures, which led to the create robust models from few or noisy samples and improved the accuracy and processing speed of the neural networks.



ANIL PAHWA (Life Fellow, IEEE) received the B.E. degree (Hons.) in electrical engineering from the Birla Institute of Technology and Science, Pilani, India, in 1975, the M.S. degree in electrical engineering from the University of Maine, Orono, ME, USA, in 1979, and the Ph.D. degree in electrical engineering from Texas A&M University at College Station, College Station, TX, USA, in 1983. Since 1983, he has been with Kansas State University, Manhattan, KS, USA, where he is currently a University Distinguished Professor and holds the Logan Fetterhoof Chair with the Department of Electrical and Computer Engineering. The National Academy selected him as a Jefferson Science Fellow, in 2014, to serve as a Senior Science Advisor with the U.S. State Department for one year. He was with the East Asian and Pacific Affairs Bureau on international policies to facilitate higher deployment of renewable energy. His research interests include distribution automation, distribution planning, renewable energy integration into power systems, and intelligent computational methods for distribution system applications.