IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

IEEE RELIABILITY SOCIETY SECTION

Received 5 August 2025, accepted 19 August 2025, date of publication 22 August 2025, date of current version 9 September 2025.

*Digital Object Identifier 10.1109/ACCESS.2025.3601999*

RESEARCH ARTICLE

# Machine Learning Framework for Industrial Machine Sound Classification in Predictive Maintenance

**NUR FATINAH MD HAFIZ[1], SYAMSIAH MASHOHOR[1,2], (Member, IEEE), MOHAMMAD HABIB SHAH ERSHAD MOHD AZRUL SHAZRIL[1], AZIZI MOHD ALI[1,2], AND MOHD FADLEE A. RASID[1,2]**

[1]Department of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia, Serdang, Selangor 43400, Malaysia
[2]Wireless and Photonics Networks Research Centre (WiPNET), Faculty of Engineering, Universiti Putra Malaysia, Serdang, Selangor 43400, Malaysia

Corresponding author: Syamsiah Mashohor (syamsiah@upm.edu.my)

This work was supported by Universiti Putra Malaysia through an Incentive Grant 9795500.

**ABSTRACT** Predictive maintenance, utilising anomalous sound classification, demonstrates a strong potential to identify mechanical faults in industrial machinery. This research proposes a machine learning-based framework for classifying anomalous sounds in industrial machines, with a particular focus on CT scan machines and fan units. The study utilises both real-world data from CT scan machine sound and the Malfunctioning Industrial Machine Investigation and Inspection (MIMII) dataset. It offers a comprehensive analysis of sound signal processing techniques, synthetic data generation methods, feature extraction processes, and classification using machine learning models to support predictive maintenance applications. In this research, sound data from a CT scan machine was collected using an Internet of Things (IoT) connected microphone located on the machine in a Klang Valley hospital. Due to the limited availability of faulty condition data, synthetic anomalous data for both operational and non-operational conditions were generated using a noise injection method. Features derived from Mel Frequency Cepstral Coefficients (MFCCs) and Mel Spectrogram representations were employed to analyse the sound data. The dataset for CT scan machine sounds is categorised into four distinct classes: anomalous operational sound (Aop), anomalous non-operational sound (Anop), normal operational sound (Nop), and normal non-operational sound (Nnop). In contrast, the MIMII dataset is classified into two categories: normal and abnormal. A Convolutional Neural Network (CNN) model was used for a sound classification system, achieving training accuracies of 98.22% with Mel spectrogram features and 98.12% with MFCC features. The results emphasise the possibility of using CNN-based sound classification to effectively anticipate and maintain CT scan machines. This finding also has the potential to be applied to predictive maintenance applications by detecting both normal and anomalous operating sounds in industrial machinery.

**INDEX TERMS** Artificial intelligence, predictive maintenance, sound signal processing, industrial machine, machine learning, Mel frequency cepstral coefficient, Mel spectrogram, convolutional neural network.

## I. INTRODUCTION

In the modern era, the healthcare industry is heavily dependent on various types of medical equipment to assist in disease diagnosis, patient monitoring, and rehabilitation. As one of the fastest-growing global sectors, healthcare demands advanced and reliable medical technology [1], [2]. However, exposure to machine failures can pose safety risks and quality issues in the machinery industry [3]. The research by [4] highlights the importance of maintaining medical equipment in addressing issues such as significant damage and prolonged downtime caused by long-standing problems. Effective maintenance management is crucial to reducing industrial device failures and addressing operational issues related to medical equipment [5].

A notable example is the CT scan machine, which employs advanced imaging technology to produce detailed cross-sectional images of the body. The data acquisition

The associate editor coordinating the review of this manuscript and approving it for publication was Haidong Shao.

system utilises an X-ray tube to measure radiation attenuation as the X-rays pass through the patient, enabling the production of accurate diagnostic images [6]. The X-ray tube of CT equipment produces X-rays to generate images [7]. This process enables the generation of precise images for diagnostic purposes, ensuring accurate and reliable medical evaluations. However, studies in South African hospitals indicate that CT machines often exceed recommended usage limits, leading to frequent malfunctions as shown in Figure 1 [8]. On average, a CT scanner experiences at least 10 breakdowns per year, which underscores the need for predictive maintenance (PdM) systems to reduce maintenance costs and ensure patient safety [9].

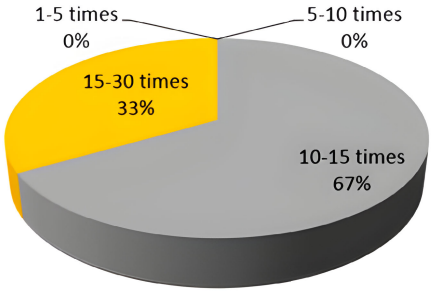



**FIGURE 1. Number of times the CT scan machine breaks down per year [8].**

The reliability of industrial machinery is critical for ensuring operational efficiency and minimising downtime in modern production systems [10]. As equipment complexity increases, predictive maintenance (PdM) has emerged as a powerful approach, leveraging various data-driven methods to anticipate equipment malfunctions before they occur. Among these methods, sound classification has gained prominence as a non-invasive and efficient technique for diagnosing the health of industrial machines by analysing their acoustic signals. The Malfunctioning Industrial Machine Investigation and Inspection (MIMII) dataset is a well-known benchmark for evaluating sound-based predictive maintenance systems for industrial machinery. It contains diverse recordings of normal and abnormal sounds from industrial machinery such as fans, valves, pumps, and sliders, providing a rich resource for the development and testing of machine learning models [11], [12]. In the research by [13], sound classification enables the identification of deviations from normal sound patterns, which often indicate mechanical anomalies or faults. Deep learning approaches, particularly Convolutional Neural Networks (CNNs), have shown remarkable performance on the MIMII dataset. The research by [14] explored the use of Continuous Wavelet Transform (CWT) with a tailored CNN architecture for anomaly detection in industrial machines.

An advanced approach involves employing an anomalous sound classification, in which the sound emitted by the machinery is monitored to identify abnormal patterns that indicate potential failure [15], [16]. Integrating advanced sound sensors with deep learning techniques can prevent premature equipment replacement and enhance maintenance efficiency [17]. By replacing manual sound-based diagnostics with automated systems, PdM improves accuracy and reduces reliance on skilled technicians. In anomalous sound detection architecture, the information of sound features is used to train machine learning models to recognize normal operating sounds and detect anomalies [18]. Existing studies on predictive maintenance and anomalous detection often rely on limited real anomalous data, leading to biased models and reduced generalisability. To address this gap, this research proposes a novel approach that develops anomalous sound detection in CT scan machines by integrating synthetic data generation, advanced feature extraction, and deep learning techniques. This comprehensive methodology enhances the classification performance of CT scan machine sounds across four distinct dataset classes. The proposed method incorporates synthetic anomalous sound generation using noise injection, effectively creating a more balanced and representative dataset for improved model training and evaluation. The acoustic signals from the CT scan machine were extracted using Mel Frequency Cepstral Coefficients (MFCCs) and Mel spectrograms, which served as input features for the models. The proposed approach leverages Convolutional Neural Networks (CNNs) to effectively capture spatial patterns within Mel spectrograms and MFCC features. The performance of the CNN classifier was validated using the MIMII dataset. This paper aims to contribute to the development of robust and reliable PdM systems capable of minimizing equipment failures and optimising industrial operations.

This study emphasizes several distinctive aspects that distinguish it from previous work. First, the synthetic data generation approach goes beyond conventional methods by embedding actual broken machine acoustic signals into real CT scan operational recordings, rather than simply introducing white noise or artificial distortions. This technique ensures a higher degree of realism and preserves domain-specific fault patterns that are critical for reliable classification. Second, the integration of IoT-based sensor systems in a medical environment adds further novelty. Unlike generic industrial monitoring, the deployment in a healthcare-critical context on CT scan machines introduces challenges such as stringent acoustic isolation, precise data acquisition, and safety-critical operational constraints. Lastly, the study contributes to the field by applying sound-based predictive maintenance to a rarely explored domain, medical imaging equipment, demonstrating the practical feasibility and potential of machine learning to improve reliability in clinical diagnostics.

### A. PREDICTIVE MAINTENANCE IN INDUSTRIAL EQUIPMENT

The advent of Industry 4.0 has resulted in the extensive use of intelligent systems, machine learning, and PdM methodologies across several industries [19]. PdM has

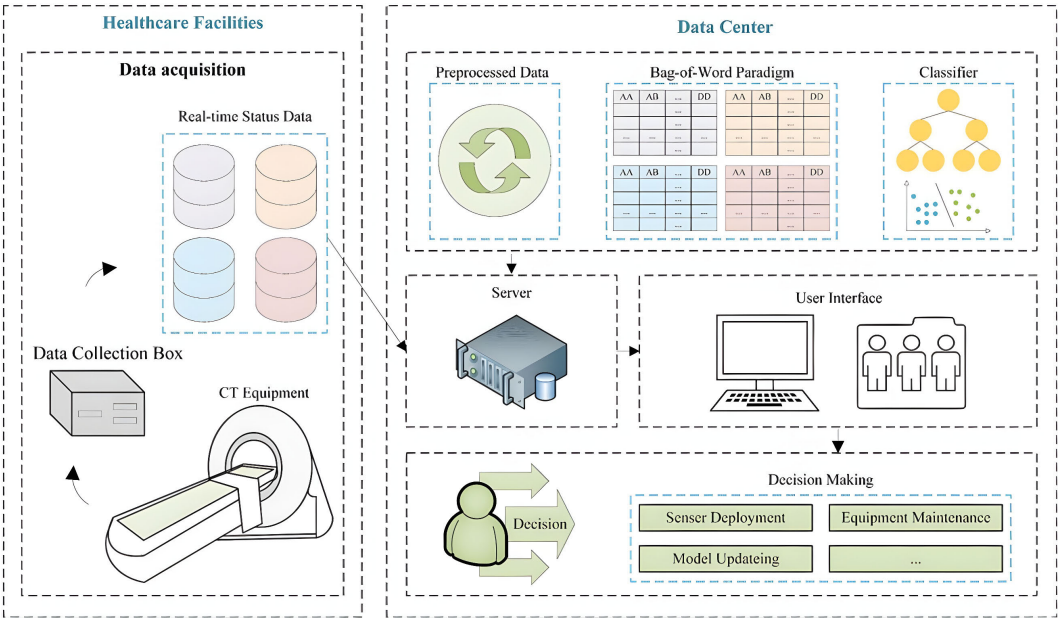



**FIGURE 2.** Real-time decision-making architecture of the system [26].

become an important approach in the modern machinery industry to maintain the reliability and health performance of a machine. Predictive maintenance involves continuous monitoring of equipment performance and condition during routine operations in order to minimise the probability of a breakdown [20]. The complexity of machinery has led to the development of predictive maintenance models that evaluate the risk of industrial equipment failure based on data analysis of daily usage and the machine's service life [20]. In [21], it was mentioned that the implementation of PdM in the medical equipment industry can help to make optimal decisions to ensure the equipment's operation by continuously monitoring its real-time performance using vast data streams.

As the industry evolves, the integration of innovative technologies, such as predictive maintenance using IoT sensors and artificial intelligence (AI), becomes crucial in ensuring the optimal performance of medical devices and sustaining the overall quality of healthcare services [22], [23]. The implementation of predictive maintenance has recently been extended to medical equipment, underscoring its critical role in the healthcare industry [24]. According to the research in [25], the authors introduce an innovative multivariate time-series classification approach that utilises the status data obtained from the Internet of Medical Things (IoMT) to forecast abnormalities in CT equipment. The predictive maintenance system proposed by [26] is based on a decision-making system that utilises tube scanning time, electrical energy consumption, the number of arcs per day and real-time IoT data, including oil temperature sensors, and voltage sensors on CT equipment, to detect equipment anomalies as shown in Figure 2. The proposed system includes data collection, data pre-processing, language processing, feature extraction, and a classification model with parameter selection.

### B. DATA ANALYSIS AND RELIABILITY

Understanding the characteristics of the acoustic signals emitted by a CT scan machine is fundamental for robust sound-based predictive maintenance systems. Signal features can be broadly classified into the time, frequency, and time-frequency domains, each contributing unique analytical advantages. The time domain captures raw amplitude variations over time, directly reflecting sound pressure fluctuations. Although this provides an immediate depiction of transient events, it suffers from high dimensionality and susceptibility to environmental noise, reducing its reliability for complex pattern recognition tasks unless further processed [27]. The frequency domain, typically represented using the Discrete Fourier Transform (DFT) or power spectrum, provides information on the distribution of spectral components such as harmonics and resonant frequencies. This domain is particularly useful for identifying stationary sound patterns characteristic of specific machine states. Frequency analysis also allows the model to isolate relevant frequency bands, improving reliability and reducing sensitivity to ambient noise from the broadband [28].

The time-frequency domain, illustrated by Mel spectrograms and Mel Frequency Cepstral Coefficients (MFCCs), synthesizes both temporal and spectral information. This dual representation allows for the detection of transient changes in the spectral content, which is especially important in identifying operational state shifts in CT scan machinery. MFCCs, derived from the Mel spectrum, compress this information using perceptual filtering and decorrelation, making them robust to background noise and dimensionality

challenges [29], [30]. Studies have shown that Mel spectrograms and MFCCs outperform traditional time or frequency domain features in audio classification, offering improved robustness and generalization under varying acoustic conditions [31]. Data preprocessing techniques such as denoising, signal-to-noise ratio (SNR) enhancement, and dimensionality reduction can lead to partial loss of original audio information. Nevertheless, these operations are essential for boosting the robustness, computational efficiency, and interpretability of the model. Their primary function is to filter out irrelevant noise, highlight discriminative features, and simplify the input for more effective learning [32].

### C. FEATURE EXTRACTION METHODS FOR SOUND SIGNAL PROCESSING

Feature extraction on sound signals is employed to convert and extract the sound data information into a format that by the machine learning models can comprehend. Feature extraction is an important process for the implementation of classification, recognition, and prediction algorithms [33], [34], [35]. Commonly used methods for feature extraction includes wavelet analysis, chroma, cepstral domain features, image-based features, and deep features [36]. For instance, in [37] paper, a snoring sound classification system was proposed using multi-feature extraction techniques like short-time Fourier transform (STFT), root mean square (RMS), spectral centroid, bandwidth, RollOff, zero-crossing rate (ZCR), and Mel Frequency Cepstral Coefficients (MFCC) achieving 99.7% accuracy with a CNN classifier. Additionally, wavelet-based features and statistical techniques have been extensively utilised in fault diagnosis for CNC machine tools, while feature extraction methods such as the Fast Fourier Transform (FFT) and wavelet transform have been effectively applied in automotive fault diagnosis [38], [39]. This process involves transforming of raw audio signals into a set of measurable and informative parameters that encapsulate the essential attributes of the sound. The studies by [40] and [41] discussed the integration of MFCC and Mel spectrogram feature extraction techniques applied in industrial equipment sound analysis. For example, Figure 3 illustrates the image representation of Mel spectrogram features used in pump fault detection within industrial machinery.

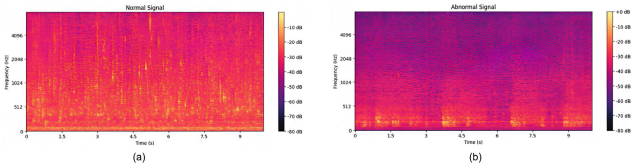



**FIGURE 3.** Mel spectrogram of (a) normal and (b) abnormal sound [34].

### D. SOUND CLASSIFICATION USING MACHINE LEARNING

Sound classification in machine learning involves the ability to categorise different sounds and audio events within a sound clip. It has a wide range of practical applications, such as identifying a song's genre based on its beat or rhythm, converting spoken language into text through speech recognition, detecting emotions in a person's voice, or distinguishing speakers based on unique vocal characteristics [42], [43], [44]. Sound classification is a rapidly advancing field, with developments in technology and machine learning models driving its progress. The application of machine learning to sound classification holds immense potential for addressing practical challenges in real-world scenarios. For instance, in [14] research, a novel CNN architecture utilising continuous wavelet transform (CWT) for feature extraction achieved a remarkable test accuracy of 99.53% on the MIMII dataset, outperforming other benchmark CNN architectures such as DenseNet, EfficientNet, and VGGNet. In another study, a CNN architecture combined with Mel Frequency Cepstral Coefficients (MFCCs) has been shown to reach a classification accuracy of 99.77% in audio signal recognition, further underscoring their efficacy in sound classification tasks [45].

While traditional machine learning models such as Support Vector Machines (SVM) and Random Forests have been explored, they generally achieve lower accuracy compared to deep learning approaches. This performance gap becomes particularly pronounced when dealing with complex or high-dimensional datasets, including those involving time-series sensor data or unstructured inputs [46]. Specifically, SVM and Random Forest both achieve approximately 81.32% accuracy in motor sound classification [47], which is often insufficient for applications requiring high precision. In contrast, deep learning architectures such as Convolutional Neural Networks (CNNs) and Gated Recurrent Units (GRUs) are capable of automatically extracting hierarchical and temporal features from raw data, thereby enhancing accuracy and robustness without requiring extensive manual feature engineering. For example, a comparative study by [48] demonstrated that GRU and CNN models significantly outperformed traditional algorithms like Random Forest and SVM in anomaly detection tasks for HVAC systems, confirming the superior performance of deep learning in complex real-world scenarios. The preference for deep learning models, particularly CNNs, is further reinforced by their ability to handle complex data representations, adaptability to various datasets, and compatibility with advanced feature extraction techniques. CNNs are also highly scalable, making them suitable for a wide range of industrial sound classification tasks, from detecting mechanical faults to monitoring operational states in noisy environments [49].

Traditional machine learning models, particularly those based on hidden Markov models (HMMs), exhibit significant limitations in handling industrial acoustic data under variable noise conditions, primarily due to their sensitivity to discrepancies between training and testing environments. This sensitivity leads to a marked decline in recognition performance when faced with diverse acoustic conditions, as evidenced by the challenges in noise reduction techniques like spectral subtraction and HMM retraining [50]. Additionally, the

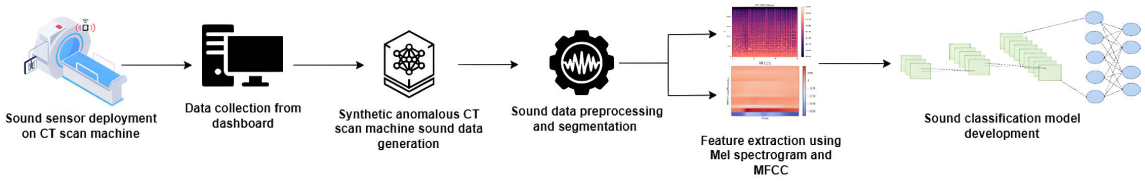



**FIGURE 4.** Overall workflow of the sound classification system.

presence of noise can exacerbate issues such as under-fitting and over-fitting, particularly in high-dimensional feature spaces, complicating the extraction of relevant information from noisy data [51]. Furthermore, the lack of robust datasets for industrial sound analysis limits the development of effective models, as demonstrated by the high sensitivity of neural network-based systems to changes in recording setups [52]. Ultimately, these limitations hinder the overall system performance, necessitating advancements in model robustness and feature selection to improve generalization capabilities in noisy environments [53].

Other than that, rule-based systems often exhibit limitations in accurately interpreting industrial acoustic signals due to the ubiquitous background noise commonly found in factory settings. This environmental noise can obscure important auditory cues essential for reliable anomaly detection. Conventional techniques, such as fixed threshold mechanisms or basic filtering methods, are typically inadequate in adapting to the dynamic and non-stationary characteristics of industrial noise, resulting in elevated rates of false alarms and missed fault detections [54]. In contrast, data-driven approaches, particularly those employing deep learning models such as Generative Adversarial Networks (GANs) and one-class Support Vector Machines (SVMs), have demonstrated superior capabilities in reconstructing and analyzing noisy audio signals, thereby enhancing the accuracy of anomaly identification [55], [56]. Additionally, the application of advanced filtering algorithms, including Butterworth filters, has proven effective in isolating high-frequency acoustic fluctuations that are frequently masked by ambient noise. These challenges underscore the critical need for robust sound analysis frameworks, as the failure to accurately detect anomalies can lead to overlooked equipment malfunctions, prolonged operational downtime, and increased maintenance expenditures [57].

This paper provides a comprehensive analysis of predictive maintenance, sound signal feature extraction, and machine learning-based sound classification. It emphasises the importance of predictive maintenance in industrial equipment, particularly in medical devices such as CT scan machines. The predictive maintenance system for the CT scan machine focuses on detecting anomalous sounds emitted from a faulty fan. Additionally, it reviews various machine learning approaches, highlighting the superiority of deep learning models, particularly CNNs, in achieving high accuracy for sound-based anomaly detection. The proposed CNN model architecture was validated using the MIMII dataset. The proposed method enhances classification performance across multiple dataset classes, contributing to more effective predictive maintenance solutions in industrial and healthcare settings.

## II. METHODOLOGY

The methodology employed in this study encompasses several critical stages, as illustrated in Figure 4 which contributes to the development of an effective predictive maintenance system. This structured workflow exemplifies a comprehensive and systematic approach by integrating IoT technologies for real-time data acquisition, synthetic data generation techniques to address the limitations of faulty condition data, features extraction to represent sound information and advanced machine learning algorithms for sound classification. In the end, the sound will be classified into four groups (normal operational sound, normal non-operational sound, anomalous operational sound, anomalous non-operational sound) for CT scan machine sounds, and two groups (normal and abnormal) for the MIMII dataset.

### A. DATA ACQUISITION

The process begins with the deployment of the Hikvision DS-2FP2020, a high-sensitivity condenser microphone sensor equipped with noise cancellation capabilities. This sensor was securely mounted on the gantry of the CT scan machine at a hospital located in Klang Valley, as shown in Figure 5. The strategic placement and design of the microphone effectively minimized environmental noise interference. Moreover, the CT scan machine generates a high-intensity operational sound, which naturally suppresses surrounding background noises, thereby ensuring the capture of clear and distinguishable acoustic signals for classification purposes. The microphone was employed to record audio signals under both operational and non-operational conditions. The recorded data were subsequently transmitted to a centralized dashboard, enabling real-time monitoring and continuous data collection. The sound recordings of the CT scan machine were captured at 5-minute intervals per sound file and logged into the server every 10 minutes. Data collection for normal operational and non-operational conditions commenced in February 2023. Conversely, the collection of anomalous sound data, representing both operational and non-operational states, began in October 2023, resulting in an imbalanced dataset. To mitigate this imbalance, a supplementary online dataset of Broken Machine Sound (BMS) effects was sourced from the POND5 website (www.pond5.com) and used to generate synthetic anomalous data for subsequent analysis.

The operational and non-operational sounds are cross-checked against a scanning log obtained from the hospital's radiography records. Every patient's scan is recorded with their patient ID, the start time, and the end time of the scanning session. In this context, "operational" refers to the CT scan machine actively performing scanning processes, whereas "non-operational" denotes the machine being in an idle state without any scanning activity. The anomalous operational and non-operational sounds of the CT scan machine can be further validated using maintenance reports provided by the vendor, dated December 8, 2023, and February 7, 2024. The collection of baseline sound data during regular operation is critical, as it provides a reference point for identifying and differentiating anomalous sounds. Each class consists of 1,500 samples of CT scan machine sound files, with a sampling rate of 44.1 kHz and a 32-bit depth for each sound sample.


**FIGURE 5.** Setup of the microphone sensor on the CT scan machine.

### B. SYNTHETIC ANOMALOUS SOUND DATA GENERATION

Acquiring sufficient faulty data is a significant challenge due to the high reliability of modern machines. To address this limitation, this study employs a synthetic data generation technique in which noise characteristics of malfunctioning or defective machines are synthetically injected into normal operational and non-operational sound recordings. Figure 6 shows the workflow for generating anomalous operational (Aop) and anomalous non-operational (Anop) CT scan machine sound. This is implemented using a Python-based function, "mix_and_render," designed to combine a primary audio file with the BMS dataset stored in a specified folder and render the resulting mixed output.

Audio mixing, as described by [58], [59] typically involves summing the amplitudes of multiple audio signals at each time step (sample) of the digital audio signal. In this project, a simplified form of audio mixing is applied, as represented in Equation (1), where M(t) denotes the mixed signal at time (t), A(t) represents the amplitude of the first audio signal at time (t) and B(t) denotes the amplitude of the second audio signal at time (t). X

$$M(t) = A(t) + B(t) \tag{1}$$

This approach aims to simulate realistic anomalous sound conditions that are infrequently encountered during normal operations but are essential for developing robust machine learning models capable of early fault detection and failure prediction.

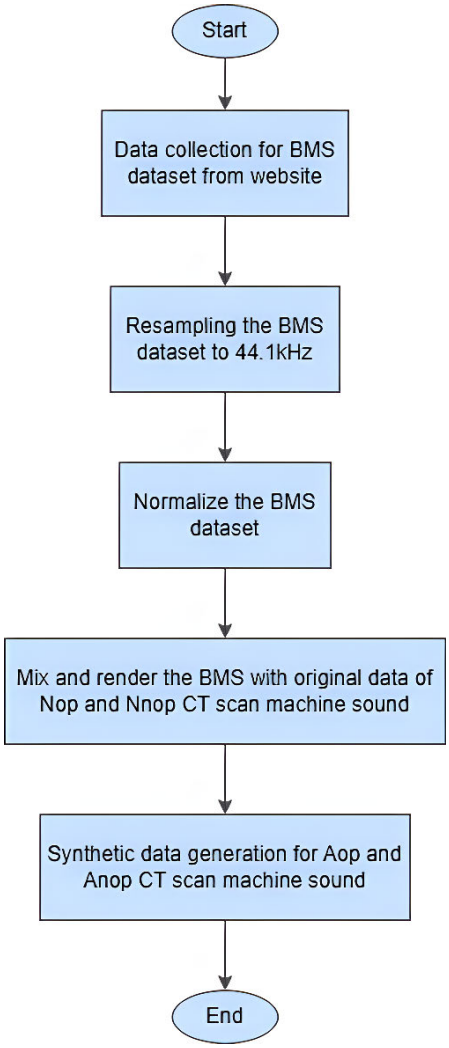



**FIGURE 6.** Workflow of synthetic data generation.

Five types of sounds were selected, including a broken fan sound, an industrial machine sound, a ventilator sound, a ventilation fan sound, and a machine whistle sound. These sounds simulate potential faults in CT scan machines, reflecting diverse failure scenarios. The BMS dataset was chosen because the CT scan machine gantry includes a cooling system with a fan component. After data collection from the BMS dataset, the BMS data is resampled from 48.0 kHz to 44.1 kHz to align with sound format of the CT scan machine. Normalisation is then applied to adjust the highest amplitude peak of the sound signals to a target value (typically $\pm 1.0$), ensuring consistent loudness without introducing clipping. The normalized BMS sounds are mixed and rendered with normal CT scan machine sounds, overlaying the fault signatures onto operational and non-operational sound data. The injected fault signals mimic

the acoustic patterns commonly associated with specific failures, such as bearing wear, motor failure, or misalignment. These artificial anomalies diversify the dataset, enriching it with a wide range of potential failure scenarios crucial for training machine learning models for anomaly detection and predictive maintenance. Synthetic anomalous sound data are compared with real anomalous data to analyse and visualise signal characteristics, improving the model's ability to identify and predict faults effectively.

Conventional synthetic data generation techniques in audio-based anomaly detection often rely on injecting white noise or artificial signal perturbations into clean datasets to simulate environmental disturbances. Although these methods have shown effectiveness in general purpose environmental sound analysis, conventional synthetic data generation techniques lack the specificity required to mimic the nuanced acoustic characteristics of actual machine faults. In contrast, this study introduces a domain-adapted approach by integrating real broken-machine sounds into CT scan operational recordings. This technique ensures that synthetic anomalies closely resemble authentic fault conditions rather than generic noise artifacts. The use of targeted fault audio, such as fan failures and machine whistles, enriches the dataset with contextually relevant variations, improving the classifier's ability to detect subtle and realistic fault signatures. This methodology not only increases the diversity of failure cases, but also enhances the model's generalization capability in high-reliability settings where faulty data are scarce. In addition, this approach represents a novel application of sound synthesis in the healthcare domain, particularly for critical diagnostic equipment such as CT scan machines, where early fault detection is essential for patient safety and system uptime.

### C. DATA PRE-PROCESSING AND SEGMENTATION

Ensuring the accuracy and reliability of sound-based machine learning models requires comprehensive data pre-processing and segmentation. These fundamental steps enhance the quality, consistency, and structure of raw sound recordings, making them suitable for the efficient extraction and classification of features. The preprocessing phase begins with visualising recorded sound signals at a 32-bit resolution and a 44.1 kHz sampling rate using Audacity software to ensure high-quality audio representation. The sound data are then segmented into distinct time frames corresponding to various CT scan machine operational states, with each segment annotated to indicate whether it represents normal or anomalous conditions. Following [60],continuous audio recordings are divided into 10-second frames extracted from 5-minute recordings, facilitating an accurate classification of machine state. Each segment is subjected to separate analysis, with the frequency content, amplitude, and spectral characteristics are extracted before feature extraction. For example, Figure 7 presents waveform plots of audio recordings obtained from an industrial CT scan machine, and Figure 8 represents waveform plots of a fan sound from MIMII dataset. Both waveforms display amplitude variations over a 10-second duration, with the x axis representing time (in seconds) and the y axis indicating amplitude (normalized between -1.0 and 1.0). These visualizations are generated as part of the preprocessing stage in sound-based anomaly detection systems. These waveforms support data labeling and quality assessment during the dataset preparation phase. Overall, the image exemplifies how waveform visualization plays a crucial role in the early stages of sound-based predictive maintenance, aiding in fault detection and improving model performance by ensuring that only meaningful sound patterns are analyzed further.

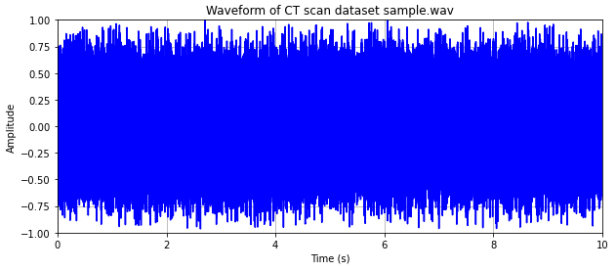



**FIGURE 7.** Waveform of normal operational CT scan machine sound.

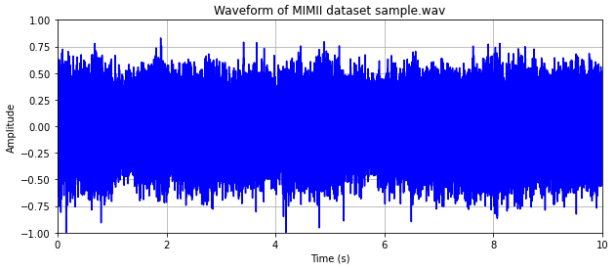



**FIGURE 8.** Waveform of normal operational fan sound of MIMII dataset.

### D. FEATURE EXTRACTION USING MEL SPECTROGRAM AND MFCC

The proposed methodology is implemented using two feature extraction techniques, Mel Spectrogram and Mel-Frequency Cepstral Coefficients (MFCC). These techniques are widely used in audio signal processing to extract meaningful features from sound data for classification and anomaly detection. The Mel spectrogram serves as a crucial tool for feature extraction, converting CT scan machine sound signals into a time-frequency representation that captures energy distribution across Mel-scaled frequency bands. The feature extraction process begins with collecting raw sound, which undergoes framing and windowing to minimise discontinuities. The Short-Time Fourier Transform (STFT) then converts the time-domain signal into a frequency-domain spectrogram, visualising frequency variations over time. A Mel filter bank is applied, mapping frequencies to the Mel scale using Equation (2), where m represents the number

of Mels units and f represents frequency in Hertz.

$$m = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \tag{2}$$

The resulting 128 × 128 Mel spectrogram represents time on the x-axis, frequency on the y-axis, and colour intensity corresponding to amplitude. In a Mel spectrogram, colour intensity represents the loudness of a specific sound frequency at any given time. Darker shades (purples/blues) signify lower amplitude sounds, while brighter colours (reds/oranges) indicate higher energy levels, with a decibel scale ranging from -80 to 0 dB. Figure 9 illustrates the Mel spectrograms for normal operational (Nop) and non-operational (Nnop) CT scan machine sounds. The energy distribution is more uniform, with fewer distinct bands, indicating stable machine operation. Orange regions suggest areas of higher energy, spread evenly across frequencies, reflecting normal conditions without significant mechanical changes. Conversely, Figure 10 presents Mel spectrograms for anomalous operational (Aop) and non-operational (Anop) sounds, showing distinct energy bands in the mid to high frequencies (approximately 2000 Hz to 8000 Hz). These concentrated bands suggest specific mechanical components generating sounds at defined frequencies, indicative of potential faults in the CT scan machine.

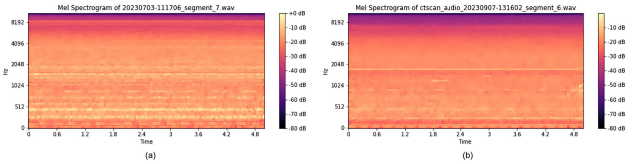



**FIGURE 9.** Mel spectrogram of (a) Nop and (b) Nnop CT scan machine sounds.

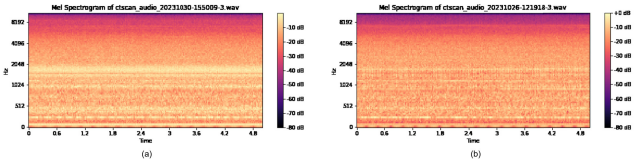



**FIGURE 10.** Mel spectrogram of (a) Aop and (b) Anop CT scan machine sounds.

Next, the Mel Frequency Cepstral Coefficients (MFCCs) are computed for each sample to extract compact and meaningful features from sound signals. MFCCs condense information into a finite set of coefficients, leveraging the auditory perception of the human ear for efficient representation. The MFCC process begins by collecting the raw sound signals and segmenting the continuous sound signal into frames due to its non-stationary nature. Windowing is then applied to minimise edge discontinuities, followed by the Discrete Fourier Transform (DFT) to convert the time-domain signal into the frequency domain. The resulting spectrum is passed through a Mel-frequency filter bank, mapping the linear frequency spectrum onto the Mel scale to emphasise perceptually relevant audio features. The final MFCCs are obtained by computing the logarithm of the filter bank energies, followed by the Discrete Cosine Transform (DCT) to decorrelate the coefficients. In this study, the Fast Fourier Transform (FFT) is configured with 2048 intervals, a sliding window of 512 points, and 25 extracted coefficients.

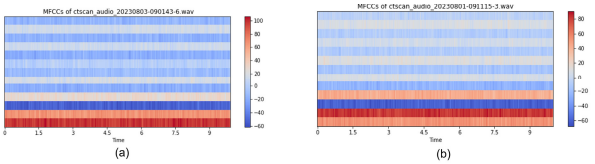



**FIGURE 11.** MFCC of (a) Nop and (b) Nnop CT scan machine sounds.

Computation is performed using the Librosa library. The MFCC plot's color scale represents signal intensity in decibels (dB), with red shades indicating higher energy (strong frequency components) and blue shades representing lower magnitudes. Figure 11 presents MFCC representations for normal operational (Nop) and non-operational (Nnop) CT scan machine sounds. The Nop feature exhibits a higher energy concentration in the lower frequencies, typical of steady-state operational sounds. The Nnop feature exhibits a similar pattern with slight variations in intensity and energy distribution. Figure 12 illustrates MFCCs for anomalous operational (Aop) and non-operational (Anop) sounds. The Aop features highlights strong low-frequency components, while the Anop features demonstrates greater variability in MFCC magnitude across time, particularly in the mid-to-lower coefficients, suggesting irregular machine behaviour.

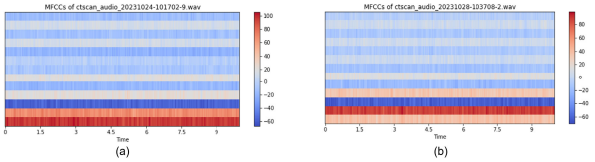



**FIGURE 12.** MFCC of (a) Aop and (b) Anop CT scan machine sounds.

### *E. DEVELOPMENT OF CT SCAN MACHINE SOUND CLASSIFICATION MODEL*

Following feature extraction, the dataset was split into training (80%) and testing (20%) sets for the development of the machine learning model. A Convolutional Neural Network (CNN) was implemented to classify normal and abnormal CT scan machine sounds, utilising Mel spectrograms and MFCC images as input. CNNs are well-suited for processing visual representations of audio, as they effectively extract spatial and hierarchical features. The convolutional layers identify frequency distributions and temporal variations crucial for distinguishing sound patterns. Figure 13 illustrates the CNN architecture used in this study. The CNN model used in this study consists of seven layers. The first is a 2D convolutional layer with 64 filters, followed by batch normalisation, ReLU activation, and max pooling. To mitigate overfitting, a dropout rate of 0.25 is applied.

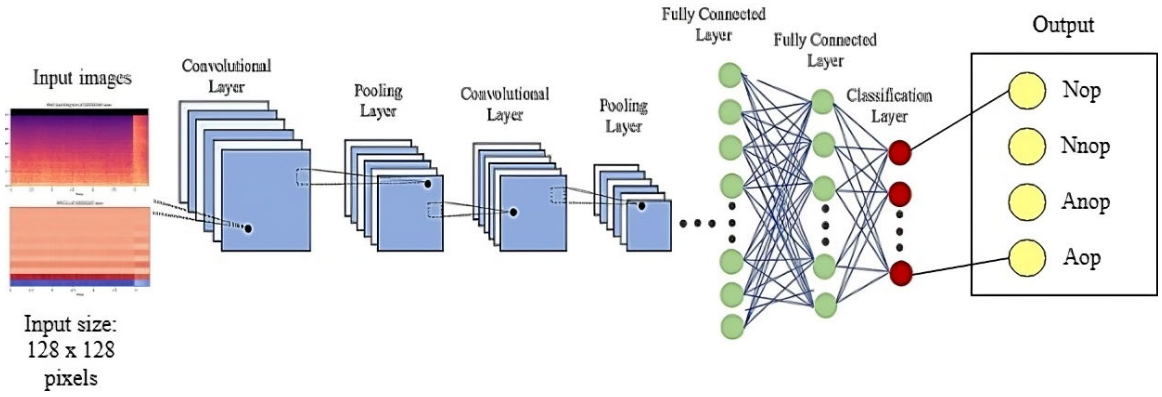



**FIGURE 13.** CNN architecture used for the proposed sound classification using Mel spectrogram or MFCC features as input.

In this study, the CNN comprises three convolutional layers with filter sizes of 3×3, increasing filter counts of 32, 64, and 128, and input Mel spectrograms of size 128×216 with a single channel. The complexity analysis confirms that the model maintains a relatively low computational footprint, making it suitable for real-time classification tasks in industrial environments where processing resources may be limited. The first and second convolutional layers follow the same structure but contain 128 and 512 filters respectively. A flattening layer transforms the data into a one-dimensional array for the fully connected layer. The model was trained over 100 epochs with 10 iterations per epoch. The CNN classifier distinguishes between normal and anomalous sounds by analysing time-frequency representations. Convolutional layers detect frequency irregularities, while pooling layers reduce dimensionality while preserving key features. Fully connected layers learn to associate patterns with normal or anomalous classes, with the final layer assigning probabilities to each category. Each input image was normalised to 128 × 128 pixels.

The machine learning framework was developed in Python using Keras with a TensorFlow backend. The implementation utilised libraries including NumPy, Pandas, Matplotlib, TensorFlow, Scikit-learn, and Seaborn. The development of a CT scan machine sound classification model integrates advanced audio processing and machine learning to enable intelligent predictive maintenance. The CNN classifier was designed to train and validate classification models using Mel spectrograms and MFCCs while assessing the impact of synthetic data across five dataset cases, as listed in Table 1. The trained CNN classifier serves as the foundation for testing, validation, and deployment, encapsulating knowledge from diverse training examples. Its generalisation ability depends on a well-designed training process, highlighting the importance of an optimized learning pipeline for accurate and reliable classification in machine learning applications.

In the implemented model, there are 4 convolutional layers, with filter counts of 32, 64, 128, and 256 respectively, and each using 3×3 filters. The input is a Mel spectrogram image of shape 128×431×1 (height, width, channels). As the network progresses deeper, the number of channels increases while the spatial dimensions decrease due to max pooling. The convolution operations dominate the computational cost, especially in deeper layers, where the number of filters is high. On actual hardware (Ryzen 7 CPU, Nvidia GTX 1650 Ti GPU), the training time to complete 100 epochs is approximately 26 minutes and 42 seconds, and the inference time (testing) for a single input takes approximately 3.053 seconds. These practical measurements align with the theoretical complexity, which scales linearly with the number of filters and quadratically with filter size. Therefore, the model remains efficient for real-time classification tasks while maintaining accuracy and feature richness.

## III. RESULTS AND DISCUSSIONS

The CNN model was trained to classify sound data, with performance metrics evaluated on both real and synthetic datasets. Optimising feature extraction parameters played a crucial role in improving CNN classification accuracy. Mel spectrograms and MFCCs emphasised different spectral and temporal characteristics, impacting model performance across five dataset cases as shown in Table 1. The combinations of training and testing data were varied to assess the effectiveness of the trained model in real or synthetic datasets. The final goal is to achieve the best model that can produce the most accurate results in recognising the real data.

**TABLE 1.** Combination of dataset cases.

| Case No | Train | Test |
|---|---|---|
| Case 1 | Synthetic data | Real data |
| Case 2 | Real data | Real data |
| Case 3 | Synthetic and Real data | Synthetic and Real data |
| Case 4 | Synthetic and Real data | Real data |
| Case 5 | Real data | Synthetic data |

Table 2 presents the testing accuracy results using Mel spectrogram and MFCC features for five combinations of training and testing data. The evaluation showed that Mel spectrograms achieved higher accuracy with n_mels = 128, while MFCCs performed best with n_mfcc = 40, highlighting the importance of parameter selection in improving

**TABLE 2.** The testing accuracy of the cnn model with different parameters using Mel spectrogram and MFCC.

| *Case* | *Mel Spectrogram* | | | *MFCC* | | |
|---|---|---|---|---|---|---|
| | *n_mels*=32 | *n_mels*=128 | *n_mels*=256 | *n_mfcc*=13 | *n_mfcc*=25 | *n_mfcc*=40 |
| 1 | 47.67 | 51.83 | 46.25 | 54.50 | 51.58 | 48.25 |
| 2 | 66.67 | 65.25 | 63.33 | 61.50 | 68.42 | 60.25 |
| 3 | 71.08 | 72.42 | 73.17 | 73.50 | 74.75 | 75.58 |
| 4 | 65.08 | 68.75 | 67.67 | 68.92 | 70.33 | 69.75 |
| 5 | 62.42 | 66.67 | 61.00 | 43.08 | 69.08 | 72.25 |
| **Average** | 63.18 | 64.98 | 62.28 | 60.30 | 66.83 | 65.21 |

classification performance. The model was tested using three different n_mels values: 32, 128 and 256. Among these, the configuration with n_mels=128 consistently yielded the highest average accuracy of 64.98%, outperforming the other two settings. This result indicates that 128 Mel frequency bands strike an optimal balance between spectral resolution and generalisation, enabling the model to capture meaningful acoustic patterns from CT scan machine sounds without introducing excessive noise or overfitting. Similarly, for MFCC features, three configurations of n_mfcc = 13, 25, and 40 were evaluated. The best performance was achieved with n_mfcc = 25, which attained an average accuracy of 66.83%. This finding suggests that increasing the number of MFCC coefficients beyond the traditional 13 coefficient enhances the model's ability to extract more detailed frequency information, which is particularly useful for detecting subtle anomalies in machine operation.

Next, the orange-highlighted values in Table 2 indicate the optimal dataset combination, determined based on testing accuracy, confusion matrices and classification reports. Among the different cases, Case 4, which utilises MFCC features, was identified as the most effective configuration for CT scan machine sound classification. It achieved high testing accuracy and superior classification performance on real data, demonstrating its effectiveness in distinguishing normal and anomalous sounds. Both the Mel spectrogram and MFCC features in Case 4 outperformed those in Cases 1 and 2, showcasing better generalisation capabilities. In particular, Case 4 achieved higher testing accuracy than Case 2, which consists of real data only. This indicates that the inclusion of synthetic anomalous data contributed to a more robust and generalisable CNN model, enhancing its ability to accurately classify real-world anomalies. Although Case 3 exhibited the highest testing accuracy, it was not selected as the optimal dataset because its test set included a mix of synthetic and real data, rather than solely evaluating performance on real-world cases. The presence of synthetic data in the test set may have artificially inflated accuracy values, preventing an accurate assessment of the model's generalisation ability. Thus, Case 4 was chosen as the ideal dataset configuration as it provided strong real-data classification performance while benefiting from synthetic anomalous data to improve model generalisation.

This highlights the importance of synthetic data augmentation in enhancing predictive maintenance models for classifying CT scan machine sounds. The integration of synthetic data enhanced model robustness while maintaining real-world relevance, establishing a reliable approach for sound classification in predictive maintenance. Despite these challenges, the CNN model with MFCC features demonstrated robust and consistent performance, making it a reliable tool for sound-based predictive maintenance.

**TABLE 3.** Accuracy results for each model using Mel Spectrogram and MFCC.

| Mode | Case | Mel Spectrogram | MFCC |
|---|---|---|---|
| **CRNN** | 1 | 69.08 | 71.08 |
| | 2 | 57.58 | 69.25 |
| | 3 | 49.83 | 72.08 |
| | 4 | 69.75 | 66.75 |
| | 5 | 73.00 | 72.67 |
| **YAMNet** | 1 | 50.25 | 52.08 |
| | 2 | 62.10 | 67.24 |
| | 3 | 65.83 | 72.43 |
| | 4 | 52.02 | 52.19 |
| | 5 | 51.65 | 51.74 |
| **LSTM** | 1 | 48.58 | 53.50 |
| | 2 | 73.25 | 65.25 |
| | 3 | 78.17 | 77.83 |
| | 4 | 69.75 | 67.58 |
| | 5 | 41.67 | 64.42 |

Additionally, several deep learning models were employed on the CT scan machine dataset to evaluate the performance of the proposed CNN model. The comparison was conducted with other state-of-the-art architectures, including the Convolutional Recurrent Neural Network (CRNN) model, Yet Another Multitask Network (YAMNet) model, and Long Short-Term Memory (LSTM) model [61], [62], [63]. Each model was trained and tested using Mel Spectrogram and MFCC feature representations to ensure a fair evaluation. The CRNN model, combines convolutional layers with recurrent layers, allowing it to capture both spatial and temporal features in the sound data. YAMNet, a pre-trained model based on the MobileNet architecture, was fine-tuned on the CT scan dataset for audio classification tasks. Meanwhile, the LSTM model utilizes sequential processing capabilities to learn time-based patterns in input audio.

The precision of each model was tested on five cases to observe consistency and convergence as shown in Table 3. The results reveal that the CRNN model consistently

outperforms the others, especially when using MFCC features, with accuracy values that peak at 73.00% Mel spectrogram and 72.67% MFCC. YAMNet shows competitive performance, especially for MFCC-based classification, achieving up to 72.08%, while the LSTM model exhibits higher variance but peaks at 78.17% (Mel) and 77.83% (MFCC), indicating strong performance in certain scenarios but reduced consistency. This comparative analysis highlights that while CRNN offers more stable and balanced performance across feature types, LSTM can yield higher peak accuracy but suffers from variability. YAMNet provides a robust pre-trained alternative with decent generalization. These findings offer insights into the strengths and limitations of each model architecture for real-world predictive maintenance applications involving audio data.

The confusion matrices and classification reports for Case 4 were analysed to evaluate the CNN classifier's performance using MFCC features. Figure 14 presents the confusion matrices for training and testing accuracy, indicating the high classification accuracy in the training dataset and effective differentiation of all four sound classes for Case 4. For the test dataset (20% unseen data), the model demonstrated strong generalisation, particularly in classifying anomalous non-operational (Anop), normal operational (Nop) and normal non-operational (Nnop) sounds. Misclassification primarily occurred between anomalous operational (Aop) and normal operational (Nop) classes. In Case 4, Aop was frequently misclassified as Nop, while the other classes were correctly classified, highlighting the challenge of distinguishing these anomalies due to their similar acoustic characteristics. These findings underscore the difficulty in accurately distinguishing between anomalous operational (Aop) and non-operational (Nop) states, which is likely attributed to the inherent similarity in their acoustic patterns and characteristics. Both classes involve the CT scan machine in an active state, resulting in overlapping spectral and temporal features that challenge the model's ability to differentiate between them. The subtle variations in sound intensity or harmonic structure may not be sufficiently captured by the current feature representation, leading the CNN to misclassify instances and generalise inaccurately between these closely related classes.

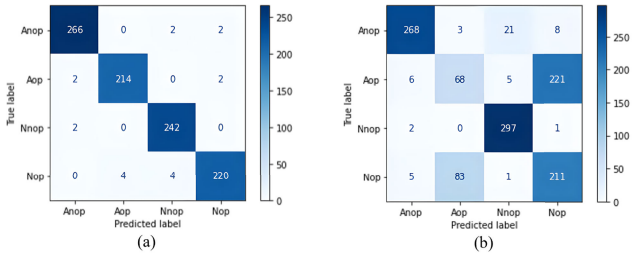



**FIGURE 14.** Confusion matrix for (a) training and (b) testing accuracy in Case 4.

In addition to that, Table 4 highlights that the model demonstrates excellent performance for the Anop class, achieving a precision of 0.95, a recall of 0.89, and an F1-score of 0.92, indicating strong reliability in detecting non-operational anomalies. Similarly, the Nnop class achieves outstanding results with a precision of 0.92, a recall of 0.99, and an F1-score of 0.95, reflecting the model's ability to identify normal idle machine sounds. In contrast, the model struggles with the Aop class, recording a low precision of 0.44, a recall of 0.23, and an F1-score of just 0.30. This suggests significant difficulty in detecting anomalous operational states, probably due to acoustic similarities with the Nop class, which also shows relatively weak performance. These results suggest that the model may struggle to distinguish between operational conditions in which faulty and normal states share overlapping spectral and temporal characteristics.

**TABLE 4.** Classification report of testing accuracy for Case 4.

| | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| **Anop** | 0.95 | 0.89 | 0.92 | 300 |
| **Aop** | 0.44 | 0.23 | 0.30 | 300 |
| **Nnop** | 0.92 | 0.99 | 0.95 | 300 |
| **Nop** | 0.48 | 0.70 | 0.57 | 300 |
| **Accuracy** | | | 0.80 | 1200 |
| **Macro avg** | 0.70 | 0.70 | 0.69 | 1200 |
| **Weighted avg** | 0.70 | 0.70 | 0.69 | 1200 |

To thoroughly evaluate the stability and generalization capability of the CNN model, a 5-fold cross-validation strategy was employed. The dataset was partitioned into five distinct folds, with each fold serving as the validation set exactly once, while the remaining four folds were used for training. This iterative process allows for a more robust estimation of the model's performance by mitigating the bias that can arise from a single train-test split. Table 5 presents the accuracy and loss achieved for each of the five folds when using MFCC as the extracted features. The results demonstrate remarkable consistency across the folds, with accuracy values that range narrowly from 0.970 (97.0%) to 0.980 (98.0%). Consequently, the loss values remained consistently low, fluctuating only between 0.020 and 0.030. The average accuracy across all five folds was calculated to be 0.978, with an average loss of 0.022. This high degree of consistency in performance metrics across different data partitions strongly indicates the CNN model's stability and its ability to generalize effectively to unseen data. The minimal variance observed suggests that the model is not overly sensitive to the specific subsets of training data and is robust in its predictive capabilities.

These results suggest that both anomalous and normal operational CT scan machine sounds contain complex patterns, making classification difficult. Performance analysis across all cases using MFCC features reveals that combining real and synthetic data (Case 3 and Case 4) improves training and testing performance, resulting in well-trained models with minimal overfitting and strong generalisation. In contrast, Case 2 exhibited validation instability, while

**TABLE 5.** Cross-validation results of MFCC-based classification model.

| Fold | Accuracy | Loss |
|---|---|---|
| 1 | 0.980 | 0.020 |
| 2 | 0.980 | 0.020 |
| 3 | 0.980 | 0.020 |
| 4 | 0.970 | 0.030 |
| 5 | 0.980 | 0.020 |
| **Average** | **0.978** | **0.022** |

Cases 1 and 5 showed overfitting, excelling in training but struggling with unseen data. These insights emphasize the need for refining data representation, enhancing feature extraction, and optimizing training strategies to improve classification performance in predictive maintenance.

To further validate the robustness of the proposed CNN based sound classifier, the MIMII dataset, comprising real sound data from normal and abnormal industrial fan operations, was used as an external benchmark. A total of 320 sound samples were used for retraining, with 80 used for testing. This assessment determined whether the CNN model, initially designed for CT scan machine sounds, maintained its performance across different datasets and operational conditions. By using the same architecture for both datasets, the study ensured consistency in evaluation, highlighting the model's adaptability to industrial sound classification. The reliance on the MIMII dataset as the sole source for evaluating the performance of the proposed CNN classifier represents a limitation in assessing the model's overall stability and robustness. Although the MIMII dataset encompasses a range of machine types and incorporates realistic industrial noise conditions, it may not fully capture the diversity and variability present in real-world environments, including differences in operational contexts, machine behaviors and anomaly patterns. To overcome this constraint and strengthen the evidence supporting the model's generalizability, future research will emphasize comprehensive cross-dataset validation. Specifically, the classifier will be evaluated using alternative publicly available datasets relevant to anomalous sound detection. This extended validation strategy will enable a deeper investigation into the model's capacity to maintain performance when exposed to previously unseen data distributions, diverse background noise levels, and varying anomaly characteristics, thus enhancing its credibility and applicability in larger industrial and real-world settings.

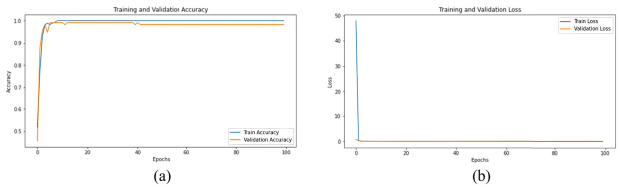



**FIGURE 15.** (a) Training and validation accuracy graph and (b) training and validation loss graph the MIMII dataset using MFCC.

Performance metrics, including accuracy, confusion matrices, classification reports, and ROC-AUC curves were analysed. Figure 15 illustrates the training and validation accuracy, along with loss curves, for the MIMII dataset using the CNN model with MFCC features. The accuracy rapidly stabilises near 100%, while the loss values decline sharply before approaching zero, indicating rapid learning with minimal errors. The confusion matrices in Figure 16 demonstrate high classification accuracy, with minor misclassification between normal and abnormal sounds. The model correctly classified 71 normal and 68 abnormal samples, with one normal sample misclassified as abnormal and four abnormal samples misclassified as normal. Misclassified cases in the testing set can be attributed to overlapping acoustic characteristics between classes, especially if the sounds share similar temporal or spectral patterns. External factors such as background noise, recording variations or subtle anomalies that resemble normal behavior could also contribute to these errors. Overall, the slight discrepancy between training and testing performance reflects a well-trained and stable model with minimal overfitting.

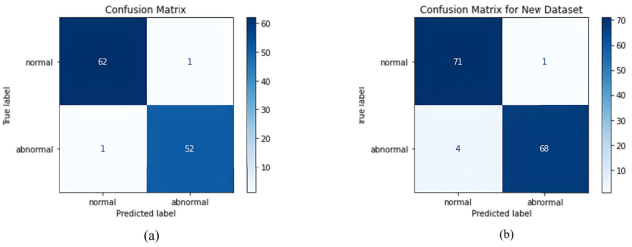


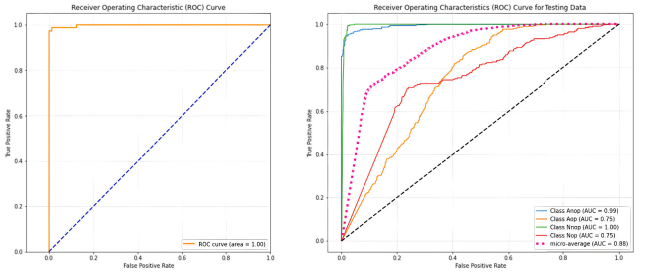

**FIGURE 16.** Confusion matrix for (a) training and (b) testing accuracy on the MIMII dataset using MFCC.

Table 6 presents the classification report, showing that the proposed CNN model achieved an accuracy of 0.97 on the MIMII dataset. The normal class performed slightly better, with a precision of 0.95, a recall of 0.99, and an F1-score of 0.97, while the abnormal class had a precision of 0.90, a recall of 0.94, and an F1-score of 0.92. The overall accuracy reached 97%, with both macro and weighted averages of precision, recall, and F1-score consistently at 0.97. This reflects a balanced performance across both classes, suggesting that the model is not biased toward any particular class. The high scores across all metrics confirm the effectiveness of the MFCC feature representation in capturing discriminative characteristics of machine sound signals, enabling accurate classification.

**TABLE 6.** Classification report of testing accuracy the MIMII dataset using MFCC.

| | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| **Normal** | 0.95 | 0.99 | 0.97 | 72 |
| **Abnormal** | 0.99 | 0.94 | 0.96 | 72 |
| **Accuracy** | | | 0.97 | 144 |
| **Macro avg** | 0.97 | 0.97 | 0.97 | 144 |
| **Weighted avg** | 0.97 | 0.97 | 0.97 | 144 |

Lastly, the ROC-AUC curves in Figure 17(a) represent a binary classification problem, where a single. ROC curve is sufficient to illustrate the model's performance. In contrast, Figure 17(b) depicts a multi-class classification problem, employing the One-vs-All (OvA) approach, which generates multiple ROC curves that allowing for a more comprehensive evaluation of the classifier's ability to distinguish between different categories. Figure 17(a) shows near-perfect discrimination, with an AUC of 1.00 for validating the CNN model on the MIMII dataset. For further validation, the ROC-AUC curve was applied to Case 4 of the CT scan machine sound dataset, as shown in Figure 17(b).




(a) (b)

**FIGURE 17. ROC AUC curve of (a) the MIMII dataset (b) the Case 4 testing dataset using MFCC.**

The model performed exceptionally well for the Anop and Nnop classes, achieving AUC values of 0.99 and 1.00, respectively. However, the Aop and Nop classes showed moderate performance with AUC values of 0.75, suggesting a potential feature overlap that affects classification. The model's overall performance, reflected by a micro-average AUC of 0.88, indicates a strong ability to generalise across all classes. This score suggests that the model effectively identifies key patterns within the dataset, which makes it suitable for practical implementation. However, moderate performance in operational sound classes highlights limitations in distinguishing complex or ambiguous features. The results of the ROC AUC curve demonstrate competitive performance compared to previous studies in sound-based machine condition monitoring. Previous research has reported AUC values ranging from 0.82 to 0.90 for CNN models employing MFCC features in mechanical diagnostics. Furthermore, studies incorporating CNN and hybrid CNN-LSTM architectures have achieved AUC scores ranging from 0.85 to 0.92 in various industrial sound classification tasks. These findings suggest that the proposed CNN model, combined with MFCC-based feature extraction, performs on par with established benchmarks. This reinforces the suitability and relevance of the model for predictive maintenance applications, particularly in monitoring the operational health of CT scan machines.

However, to better understand the limitations and potential improvements of the current framework, it is important to consider the impact of the classification structure on model performance. The model was evaluated using a four-class classification structure. Reducing the number of classes (e.g., binary classification into normal vs. abnormal) typically improves accuracy, as observed in comparative testing using the MIMII dataset. On the other hand, increasing the number of classes often leads to degraded performance due to overlapping acoustic features, particularly between classes such as abnormal during operation and normal during operation. These findings are evident in the confusion matrix, where misclassifications frequently occur between these closely related classes. To improve the robustness of future predictive maintenance systems, fuzzy logic-based classification or hierarchical decision strategies can be considered. These approaches can help model uncertainty and reduce ambiguity in borderline cases, making classification more realistic and suitable for industrial applications.

The proposed system also offers several advantages for real-time industrial fault detection and predictive maintenance of medical equipment. By leveraging lightweight audio features such as Mel spectrograms and MFCCs, the system enables rapid signal processing suitable for real-time deployment. The integration of a noise cancellation condenser microphone installed on the CT scan gantry allows precise sound acquisition while suppressing ambient interference. This ensures that the machine's high-intensity operational sounds are distinctly captured, facilitating accurate classification. In terms of decision-making, the convolutional neural network (CNN) architecture employed in this study provides robust pattern recognition capabilities by learning hierarchical representations from acoustic features. This enables consistent model response across various operational states, improving the reliability of automated fault detection. From an energy efficiency perspective, early detection through acoustic analysis prevents machine overuse, minimizes unplanned downtime, and reduces the need for energy-intensive emergency repairs, thereby supporting proactive maintenance strategies.

## IV. CONCLUSION

This study demonstrated the potential of integrating machine learning into predictive maintenance through sound-based classification of industrial machines. By leveraging Mel spectrogram and MFCC features, the system effectively transformed raw audio signals into informative representations for accurate fault detection. To address the scarcity of anomalous sound data, synthetic data generation was employed, enhancing the model's generalisation across various dataset configurations. The findings confirm that combining real and synthetic data improves classification performance and reinforces the model's applicability in real-world scenarios. Despite challenges in distinguishing acoustically similar fault conditions, the CNN-based classifier demonstrated strong generalisation capabilities, particularly in differentiating between normal non-operational and anomalous non-operational sounds. This system offers significant benefits

for industrial machine sound monitoring, enabling early detection of mechanical faults, reducing unplanned downtime and supporting proactive maintenance strategies. While the current framework has proven effective in capturing clean and reliable acoustic data, this performance is largely attributed to the strategic placement of the microphone and the dominance of the CT scan machine's operational sound. In more acoustically complex environments, where multiple overlapping sound sources may be present or where the target machine sound is less prominent, the system's robustness could be affected.

In such scenarios, additional signal processing measures, such as advanced denoising algorithms or adaptive filtering techniques, may be necessary to preserve classification accuracy. Therefore, future work should include extensive testing under diverse acoustic conditions to evaluate the adaptability and stability of the model in real-time deployment, ultimately strengthening its reliability for broader industrial applications. To further enhance the realism and adaptability of the proposed predictive maintenance framework, future work should explore the integration of fuzzy logic-based classification. Traditional classifiers like CNNs rely on crisp decision boundaries, which may struggle to accurately classify ambiguous or borderline sound events—particularly in scenarios where the acoustic characteristics of fault conditions overlap with normal operations. Fuzzy logic, on the other hand, allows for degrees of membership and can model uncertainty more effectively, making it suitable for handling the inherent variability and imprecision of real-world industrial environments. By incorporating fuzzy inference systems or hybrid models (e.g., fuzzy-CNN or fuzzy decision trees), the system could offer more nuanced classifications, leading to improved interpretability and decision-making. This approach would be especially valuable in early fault detection, where symptoms may be subtle and are not easily categorized using rigid classification rules.

## REFERENCES


[1] A. Moses and A. Sharma, "What drives human resource acquisition and retention in social enterprises? An empirical investigation in the healthcare industry in an emerging market," *J. Bus. Res.*, vol. 107, pp. 76–88, Feb. 2020, doi: 10.1016/j.jbusres.2019.07.025.

[2] K. Moons, G. Waeyenbergh, and L. Pintelon, "Measuring the logistics performance of internal hospital supply chains—A literature study," *Omega*, vol. 82, pp. 205–217, Jan. 2019, doi: 10.1016/j.omega.2018.01.007.

[3] S.-M. Kim and Y. Soo Kim, "Enhancing sound-based anomaly detection using deep denoising autoencoder," *IEEE Access*, vol. 12, pp. 84323–84332, 2024, doi: 10.1109/ACCESS.2024.3414435.

[4] S. H. Salim and S. A. Salim, "Decision-making framework for medical equipment maintenance and replacement in private hospitals," *Int. J. Sustain. Construct. Eng. Technol.*, vol. 14, no. 3, Sep. 2023, doi: 10.30880/ijscet.2023.14.03.028.

[5] C. Corciov, R. Fuior, D. Andrioi, and C. Luca. *Assessment of Medical Equipment Maintenance Management*. Operations Management and Management Science. Accessed: Jan. 13, 2025. [Online]. Available: https://www.intechopen.com/chapters/1085559

[6] E. Seeram, *Computed Tomography: Physics and Technology. A Self Assessment Guide*. Hoboken, NJ, USA: Wiley, 2022.

[7] M. A. Bashir, M. H. Sanhory, F. J. Alrasheed, A. Abdelrahman, and A. A. Abdullah, "X-ray tube arc preventation by stabilization of voltage in a dual energy CT scanner: A review study," in *Proc. Int. Conf. Comput., Control, Electr., Electron. Eng. (ICCCEEE)*, Sep. 2019, pp. 1–6, doi: 10.1109/ICCCEEE46830.2019.9071025.

[8] M. J. Pita, H. W. Pretorius, M. Pita, and J. H. Pretorius, "Evaluation of computerized tomographic scanner preventive maintenance: A case study," in *Proc. 31st Annu. Southern Afr. Inst. Ind. Eng. Conf.*, Oct. 2020, pp. 258–268.

[9] H. Zhou, Z. Li, T. Wu, C. Wang, and K. Li, "Prognostic and health management of CT equipment via a distance self-attention network using Internet of Things," *IEEE Internet Things J.*, vol. 11, no. 19, pp. 31338–31354, Oct. 2024, doi: 10.1109/JIOT.2024.3421365.

[10] J. Lee, H.-A. Kao, and S. Yang, "Service innovation and smart analytics for industry 4.0 and big data environment," *Proc. CIRP*, vol. 16, pp. 3–8, Jan. 2014, doi: 10.1016/j.procir.2014.02.001.

[11] L. Gantert, T. Zeffiro, M. Sammarco, and M. E. M. Campista, "Multiclass classification of faulty industrial machinery using sound samples," *Eng. Appl. Artif. Intell.*, vol. 136, Oct. 2024, Art. no. 108943, doi: 10.1016/j.engappai.2024.108943.

[12] H. Purohit, 2019, "MIMII dataset: Sound dataset for malfunctioning industrial machine investigation and inspection (public 1.0)," doi: 10.5281/zenodo.3384388.

[13] Y. Koizumi, S. Saito, H. Uematsu, and N. Harada, "Optimizing acoustic feature extractor for anomalous sound detection based on Neyman–Pearson lemma," in *Proc. 25th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2017, pp. 698–702, doi: 10.23919/EUSIPCO.2017.8081297.

[14] N. Sreevidya, S. S. Nathala, A. Dayal, M. S. Manikandan, J. Zhou, and L. R. Cenkeramaddi, "Classification of anomalies in industrial machines utilizing machine sounds and deep learning," in *Proc. IEEE 19th Conf. Ind. Electron. Appl. (ICIEA)*, Aug. 2024, pp. 1–6, doi: 10.1109/ICIEA61579.2024.10665175.

[15] E. Di Fiore, A. Ferraro, A. Galli, V. Moscato, and G. Sperlì, "An anomalous sound detection methodology for predictive maintenance," *Expert Syst. Appl.*, vol. 209, Dec. 2022, Art. no. 118324, doi: 10.1016/j.eswa.2022.118324.

[16] Y. Ota and M. Unoki, "Anomalous sound detection for industrial machines using acoustical features related to timbral metrics," *IEEE Access*, vol. 11, pp. 70884–70897, 2023, doi: 10.1109/ACCESS.2023.3294334.

[17] Y. Wang, Y. Zheng, Y. Zhang, Y. Xie, S. Xu, Y. Hu, and L. He, "Unsupervised anomalous sound detection for machine condition monitoring using classification-based methods," *Appl. Sci.*, vol. 11, no. 23, p. 11128, Nov. 2021, doi: 10.3390/app112311128.

[18] M. Wang, Q. Mei, X. Song, X. Liu, R. Kan, F. Yao, J. Xiong, and H. Qiu, "A machine anomalous sound detection method using the lMS spectrogram and ES-MobileNetV3 network," *Appl. Sci.*, vol. 13, no. 23, p. 12912, Dec. 2023, doi: 10.3390/app132312912.

[19] Z. M. Çınar, A. A. Nuhu, Q. Zeeshan, O. Korhan, M. Asmael, and B. Safaei, "Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0," *Sustainability*, vol. 12, no. 19, p. 8211, Oct. 2020, doi: 10.3390/su12198211.

[20] J. M. Fordal, P. Schjølberg, H. Helgetun, T. Ø. Skjermo, Y. Wang, and C. Wang, "Application of sensor data based predictive maintenance and artificial neural networks to enable industry 4.0," *Adv. Manuf.*, vol. 11, no. 2, pp. 248–263, Jun. 2023, doi: 10.1007/s40436-022-00433-x.

[21] K. Purnachand, M. Shabbeer, P. N. V. S. Rao, and C. M. Babu, "Predictive maintenance of machines and industrial equipment," in *Proc. 10th IEEE Int. Conf. Commun. Syst. Netw. Technol. (CSNT)*, Jun. 2021, pp. 318–324, doi: 10.1109/CSNT51715.2021.9509696.

[22] A. Das, "AI-enabled predictive maintenance for medical imaging equipment," in *Prosthodontics Revolution: Modern Techniques in Dental Restorations*, 2022, pp. 29–35.

[23] O. Manchadi, F.-E. Ben-Bouazza, and B. Jioudi, "Predictive maintenance in healthcare system: A survey," *IEEE Access*, vol. 11, pp. 61313–61330, 2023, doi: 10.1109/ACCESS.2023.3287490.

[24] S. Sabah, M. Moussa, and A. Shamayleh, "Predictive maintenance application in healthcare," in *Proc. Annu. Rel. Maintainability Symp. (RAMS)*, Jan. 2022, pp. 1–9, doi: 10.1109/RAMS51457.2022.9893973.

[25] C. Wang, Q. Liu, H. Zhou, T. Wu, H. Liu, J. Huang, Y. Zhuo, Z. Li, and K. Li, "Anomaly prediction of CT equipment based on IoMT data," *BMC Med. Informat. Decis. Making*, vol. 23, no. 1, p. 166, Aug. 2023, doi: 10.1186/s12911-023-02267-4.

[26] H. Zhou, Q. Liu, H. Liu, Z. Chen, Z. Li, Y. Zhuo, K. Li, C. Wang, and J. Huang, "Healthcare facilities management: A novel data-driven model for predictive maintenance of computed tomography equipment," *Artif. Intell. Med.*, vol. 149, Mar. 2024, Art. no. 102807, doi: 10.1016/j.artmed.2024.102807.

[27] G. Jombo and Y. Zhang, "Acoustic-based machine condition monitoring—Methods and challenges," *Eng*, vol. 4, no. 1, pp. 47–79, Jan. 2023, doi: 10.3390/eng4010004.

[28] S. Krishnan, "Advanced analysis of biomedical signals," in *Biomedical Signal Analysis for Connected Healthcare*, 2021, pp. 157–222.

[29] D. Kusumawati, A. A. Ilham, A. Achmad, and I. Nurtanio, "Performance analysis of feature mel frequency cepstral coefficient and short time Fourier transform input for lie detection using convolutional neural network," *Int. J. Informat. Visualizat.*, vol. 8, no. 1, p. 279, Mar. 2024, doi: 10.62527/joiv.8.1.2062.

[30] J. Perianez-Pascual, J. D. Gutiérrez, L. Escobar-Encinas, Á. Rubio-Largo, and R. Rodriguez-Echeverria, "Beyond spectrograms: Rethinking audio classification from EnCodec's latent space," *Algorithms*, vol. 18, no. 2, p. 108, Feb. 2025, doi: 10.3390/a18020108.

[31] S. M. M. Pasha, S. R. Sohag, and M. M. Ali, "Enhancing audio classification with a CNN-attention model: Robust performance and resilience against backdoor attacks," *Int. J. Comput. Appl.*, vol. 186, no. 49, pp. 26–33, Nov. 2024, doi: 10.5120/ijca2024924154.

[32] M. D. Renanti, A. Buono, K. Priandana, and S. H. Wijaya, "Evaluating noise-robustness of convolutional and recurrent neural networks for baby cry recognition," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 6, 2024, doi: 10.14569/ijacsa.2024.0150660.

[33] Q. Liu, J. Zhang, J. Liu, and Z. Yang, "Feature extraction and classification algorithm, which one is more essential? An experimental study on a specific task of vibration signal diagnosis," *Int. J. Mach. Learn. Cybern.*, vol. 13, no. 6, pp. 1685–1696, Jun. 2022, doi: 10.1007/s13042-021-01477-4.

[34] M. E. Haque, S. M. Salam, and M. S. Islam, "Human speech emotion recognition using artificial neural networks technique," in *Proc. Int. Conf. Adv. Comput., Commun., Electr., Smart Syst. (iCACCESS)*, Mar. 2024, pp. 1–6, doi: 10.1109/ICACCESS61735.2024.10499562.

[35] T. J. Fawcett, C. S. Cooper, R. J. Longenecker, and J. P. Walton, "Machine learning, waveform preprocessing and feature extraction methods for classification of acoustic startle waveforms," *MethodsX*, vol. 8, Jan. 2021, Art. no. 101166, doi: 10.1016/j.mex.2020.101166.

[36] G. Sharma, K. Umapathy, and S. Krishnan, "Trends in audio signal feature extraction methods," *Appl. Acoust.*, vol. 158, Jan. 2020, Art. no. 107020, doi: 10.1016/j.apacoust.2019.107020.

[37] T. Adesuyi, B.-M. Kim, and J. Kim, "Snoring sound classification using 1D-CNN model based on multi-feature extraction," *Int. J. FUZZY Log. Intell. Syst.*, vol. 22, no. 1, pp. 1–10, Mar. 2022, doi: 10.5391/ijfis.2022.22.1.1.

[38] S. Ding, S. Zhang, and C. Yang, "Machine tool fault classification diagnosis based on audio parameters," *Results Eng.*, vol. 19, Sep. 2023, Art. no. 101308, doi: 10.1016/j.rineng.2023.101308.

[39] Y. Hao and D. Chen, "The role of noise signal processing and feature extraction in automotive fault diagnosis," in *Proc. Int. Conf. Telecommun. Power Electron. (TELEPE)*, vol. 9, May 2024, pp. 1–6, doi: 10.1109/telepe64216.2024.00136.

[40] A. S. Bin Saharom and F. Ehara, "Comparative analysis of MFCC and mel-spectrogram features in pump fault detection using autoencoder," in *Proc. 2nd Int. Conf. Comput. Graph. Image Process. (CGIP)*, Jan. 2024, pp. 124–128, doi: 10.1109/cgip62525.2024.00030.

[41] T. Tran and J. Lundgren, "Drill fault diagnosis based on the scalogram and mel spectrogram of sound signals using artificial intelligence," *IEEE Access*, vol. 8, pp. 203655–203666, 2020, doi: 10.1109/ACCESS.2020.3036769.

[42] N. N. Wijaya, D. R. I. M. Setiadi, and A. R. Muslikh, "Music-genre classification using bidirectional long short-term memory and mel-frequency cepstral coefficients," *J. Comput. Theories Appl.*, vol. 1, no. 3, pp. 243–256, Jan. 2024, doi: 10.62411/jcta.9655.

[43] V. M. Reddy, T. Vaishnavi, and K. P. Kumar, "Speech-to-text and text-to-speech recognition using deep learning," in *Proc. 2nd Int. Conf. Edge Comput. Appl. (ICECAA)*, Jul. 2023, pp. 657–666, doi: 10.1109/icecaa58104.2023.10212222.

[44] M. B. Er, "A novel approach for classification of speech emotions based on deep and acoustic features," *IEEE Access*, vol. 8, pp. 221640–221653, 2020, doi: 10.1109/ACCESS.2020.3043201.

[45] L. Muscar, T. Telembici, and C. Rusu, "Deep learning-based sound classification algorithms for enhanced service robots audio capabilities," in *Proc. 15th Int. Conf. Commun. (COMM)*, Oct. 2024, pp. 1–6, doi: 10.1109/comm62355.2024.10741397.

[46] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, Jan. 2019.

[47] S. A. Khan, F. Ahmad Khan, A. Jamil, and A. Ali Hameed, "Interpretable motor sound classification for enhanced fault detection leveraging explainable AI," in *Proc. IEEE 3rd Int. Conf. Comput. Mach. Intell. (ICMI)*, Apr. 2024, pp. 1–10, doi: 10.1109/ICMI60790.2024.10585829.

[48] M. Haidarh, C. Mu, Y. Liu, and X. He, "Exploring traditional, deep learning and hybrid methods for hyperspectral image classification: A review," *J. Inf. Intell.*, Apr. 2025, doi: 10.1016/j.jiixd.2025.04.002.

[49] P. Priyadarshinee, "Deep learning for acoustic signal processing for industrial noise," *J. Acoust. Soc. Amer.*, vol. 148, no. 4, p. 2767, Oct. 2020, doi: 10.1121/1.5147702.

[50] E. Trentin and M. Gori, "Toward noise-tolerant acoustic models," in *Proc. 7th Eur. Conf. Speech Commun. Technol. (Eurospeech)*, Sep. 2001, pp. 889–892, doi: 10.21437/eurospeech.2001-270.

[51] A. Mariello, "Learning from noisy data through robust feature selection, ensembles and simulation-based optimizatio," Ph.D. dissertation, Univ. Trento, Trento, Italy, 2019.

[52] S. Grollmisch, J. Abesser, J. Liebetrau, and H. Lukashevich, "Sounding industry: Challenges and datasets for industrial sound analysis," in *Proc. 27th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2019, pp. 1–5, doi: 10.23919/eusipco.2019.8902941.

[53] R. Bodo, M. Bertocco, and A. Bianchi, "Impact of noise on machine learning-based condition monitoring applications: A case study," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, May 2020, pp. 1–6, doi: 10.1109/I2MTC43012.2020.9129119.

[54] Y. Liu, T. Dillon, W. Yu, W. Rahayu, and F. Mostafa, "Noise removal in the presence of significant anomalies for industrial IoT sensor data in manufacturing," *IEEE Internet Things J.*, vol. 7, no. 8, pp. 7084–7096, Aug. 2020, doi: 10.1109/JIOT.2020.2981476.

[55] Y. Tagawa, R. Maskeliūnas, and R. Damaševičius, "Acoustic anomaly detection of mechanical failures in noisy real-life factory environments," *Electronics*, vol. 10, no. 19, p. 2329, Sep. 2021, doi: 10.3390/electronics10192329.

[56] S. Kilickaya, M. Ahishali, C. Celebioglu, F. Sohrab, L. Eren, T. Ince, M. Askar, and M. Gabbouj, "Audio-based anomaly detection in industrial machines using deep one-class support vector data description," in *Proc. IEEE Symp. Comput. Intell. Eng./Cyber Phys. Syst. Companion (CIES Companion)*, Mar. 2025, pp. 1–5, doi: 10.1109/CIESCOMPANION65073.2025.11010815.

[57] P. Wißbrock, Y. Richter, D. Pelkmann, Z. Ren, and G. Palmer, "Cutting through the noise: An empirical comparison of psycho-acoustic and envelope-based features for machinery fault detection," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Rhodes Island, Greece, Jun. 2023, pp. 1–5, doi: 10.1109/ICASSP49357.2023.10095756.

[58] M. G. López, H. M. Lozano, L. P. Sánchez, and L. N. O. Moreno, "Blind source separation of audio signals using independent component analysis and wavelets," in *Proc. 21st Int. Conf. Electr. Commun. Comput.*, Feb. 2011, pp. 152–157, doi: 10.1109/CONIELECOMP.2011.5749353.

[59] R. Brice, "Let's stick together—Recording consoles," in *Music Engineering*. Amsterdam, The Netherlands: Elsevier, 2001, pp. 340–387.

[60] J. J.-W. Lim, B.-Y. Ooi, S.-Y. Liew, and W.-K. Cheng, "Comparative analysis of machine learning techniques for acoustic machine tracking under different signal durations," in *Proc. 3rd Int. Conf. Artif. Intell. Data Sci. (AiDAS)*, Sep. 2022, pp. 1–6, doi: 10.1109/aidas56890.2022.9918703.

[61] A. Bansal and N. K. Garg, "Robust technique for environmental sound classification using convolutional recurrent neural network," *Multimedia Tools Appl.*, vol. 83, no. 18, pp. 54755–54772, Dec. 2023, doi: 10.1007/s11042-023-17066-2.

[62] N. H. Valliappan, S. D. Pande, and S. Reddy Vinta, "Enhancing gun detection with transfer learning and YAMNet audio classification," *IEEE Access*, vol. 12, pp. 58940–58949, 2024, doi: 10.1109/ACCESS.2024.3392649.

[63] I. Lezhenin, N. Bogach, and E. Pyshkin, "Urban sound classification using long short-term memory neural network," in *Proc. Federated Conf. Comput. Sci. Inf. Syst. (FedCSIS)*, Leipzig, Germany, Sep. 2019, pp. 57–60, doi: 10.15439/2019F185.

**NUR FATINAH MD HAFIZ** received the Bachelor of Engineering degree in computer and communication systems from Universiti Putra Malaysia (UPM), in 2022, where she is currently pursuing the Master of Science degree in computer and embedded systems engineering. She joined a research project titled ''Internet of Things (IoT) Connected Machine Learning Based Predictive Maintenance for Computed Tomography Scanner (CT-Scan)'' as a Research Assistant under the Wireless and Photonics Networks Research Centre (WiPNET), UPM. Her research interests include sound processing and artificial intelligence (AI).

**SYAMSIAH MASHOHOR** (Member, IEEE) received the B.Eng. degree from the Department of Computer and Communication Systems Engineering, Universiti Putra Malaysia (UPM), in 2002, and the Ph.D. degree from the Department of Electronics and Electrical Engineering, The University of Edinburgh, U.K., in 2006, with a focus on ''Genetic Algorithm Based Printed Circuit Board Optical Inspection.'' She was with the Department of Computer and Communication Systems Engineering, UPM, as a Tutor, in 2002, a Lecturer, in 2006, and a Senior Lecturer, from 2009 to 2019, and currently, she is an Associate Professor. Her field of research is on artificial intelligence (AI) and image processing. Most of her publications are published in various AI-related journals, and she has been awarded several local research grants in Malaysia and involved with a few international research grants. She is actively involved with multi-disciplinary research studying AI applications in medical imaging, the IoT-based predictive maintenance, and agriculture.

**MOHAMMAD HABIB SHAH ERSHAD MOHD AZRUL SHAZRIL** received the B.Eng. degree from the Department of Computer and Communication Systems Engineering, Universiti Putra Malaysia (UPM), in 2022. Currently, he is pursuing the master's degree, focusing on predictive maintenance using the IoT sensors. His research interests include artificial intelligence (AI) and image processing. He has been actively involved in notable projects, including the detection of Harmful Algae Bloom using machine vision and predictive maintenance on CT scan machines.

**AZIZI MOHD ALI** received the Bachelor of Engineering degree in computer and communication systems from Universiti Putra Malaysia, in 2002, and the M.Sc. degree in information technology from Open University Malaysia, in 2023. He was a Senior Engineer in radio planning and network optimization with Nokia Siemens Networks, working on 2G, 3G, TETRA, and WiMAX networks. In 2008, he joined UPM as a Research Officer under the Wireless and Photonics Networks Research Centre (WiPNET), UPM. He is currently a Senior Research Officer and the Head of the Funding Unit, Research Management Centre (RMC), UPM. His current research interests include wireless sensor networks, the IoT, and wireless communication.

**MOHD FADLEE A. RASID** received the B.Sc. degree in electrical engineering from Purdue University, USA, and the Ph.D. degree in electronic and electrical engineering (mobile communications) from Loughborough University, U.K. He is currently with the Faculty of Engineering, Universiti Putra Malaysia, (UPM). He was a Research Consultant for a British Council UKIERI Project on wireless medical sensors. He was also part of French Government STIC Asia Project on ICT-ADI: Toward a human-friendly assistive environment for aging, disability & independence. He directs research activities within the Wireless Sensors Group and his work on wireless medical sensors has gained importance in health care applications involving mobile telemedicine and has had worldwide publicity, including BBC news. He led a few projects on sensor networks from the Ministry of Communications and Multimedia Malaysia as well as the Economic Planning Unit (EPU) under the Prime Minister's Department, particularly for agriculture and environmental applications. He was involved with a project under Qatar National Research Fund by Qatar Foundations on Ubiquitous Healthcare and recently with the Korean Development Institute (KDI) on Smart City.