

Received 6 September 2024, accepted 26 September 2024, date of publication 30 September 2024,
date of current version 21 October 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3471169

RESEARCH ARTICLE

Acoustic Feature Extraction and Classification Techniques for Anomaly Sound Detection in the Electronic Motor of Automotive EPS

EUN SUN YUN^{ID 1,2} AND MIN JOONG JEONG^{ID 2,3}

¹Department of High-Fidelity Model Acceleration Research, Korea Institute of Science and Technology Information (KISTI), Daejeon 34141, Republic of Korea

²Department of Data and HPC, University of Science and Technology (UST), Daejeon 34141, Republic of Korea

³Supercomputing Application Center, Korea Institute of Science and Technology Information (KISTI), Daejeon 34141, Republic of Korea

Corresponding author: Minjoong Jeong (jeong@kisti.re.kr)

This work was supported in part by the Institute of Civil Military Technology Cooperation under Grant 22-SN-BD-11, in part by the Korea Institute of Science and Technology Information Grant K24-L2-M1-C4, and in part by the National Research Council of Science and Technology Grant GTL2431-000.

ABSTRACT Anomaly sound detection (ASD) is a field of research dedicated to the early identification of mechanical failures and the facilitation of preventative maintenance to avert potential hazards and optimize performance. The capability of ASD to detect defects in mechanical equipment within this domain largely depends on the ability to uncover faults that belong to unknown categories. However, the limited labeled anomaly data available in real-world settings remains a challenge in ASD. To overcome this problem, in this study, we introduce a novel method for ASD, namely LSTM-AE with MFCC and dynamic feature maximization transformation (DFMT). The proposed method can improve normal and anomaly sound distinction in the limited anomaly data. Additionally, the unsupervised learning LSTM-AE technique enables the model to effectively learn complex sound patterns and accurately identify anomalies and deviations from normal features. In the proposed method, MFCC and DFMT are combined to generate and integrate feature maps, removing noise while retaining robustness without losing crucial features. Experimental results reveal that the proposed method achieved an accuracy of 99.2%.

INDEX TERMS Anomaly sound detection, dynamic feature maximization transformation, electronic power steering, motor diagnosis.

I. INTRODUCTION

The automotive industry is at the forefront of technological innovation, integrating advanced mechanical systems to enhance automotive safety, reliability, and performance [1]. Among these crucial systems, the electronic motor of automotive electronic power steering (EPS) not only directly influences automotive control but is also a key component in conveying steering commands of the automotive [2]. EPS represents a significant advancement in automotive steering technology, replacing the traditional hydraulic power steering systems [3], [4] with an electrically powered system. The core of EPS lies in its electronic motor, which directly assists the steering mechanism, enhancing the ability of the driver to steer the automotive with less effort [4]. The

The associate editor coordinating the review of this manuscript and approving it for publication was Donato Impedovo^{ID}.

motor is activated by the Electronic Control Unit (ECU), which determines the direction and amount of assistance based on the input of the driver and automotive dynamics [4]. The motor subsequently provides the necessary torque to the steering mechanism, minimizing the effort required to turn the wheels. Thus, the motor allows for variable assistance, adjusting the level of steering help based on automotive speed, steering wheel torque, and other driving conditions [2]. This adaptability enhances vehicle handling and safety, providing light steering at low speeds and firm steering at high speeds. Therefore, the detection of anomalies in these electronic motors becomes not only a matter of maintenance but also a crucial safety imperative [5]. The most common failure in EPS is rolling bearing failure, which is also the most common cause of failure in industrial rotating machinery. Several studies have mainly focused on developing methodologies for inner and outer ring

defects [6], [7] and ball defects [8]. Research into reliable anomaly detection continues to expand using techniques such as vibration analysis, acoustic methods, thermal imaging, current analysis, and monitoring to diagnose these faults. Developing methodologies for anomaly detection remains an area of opportunity [9].

Among the various anomaly detection methods, anomaly sound detection (ASD) [10], [11], [12], [13], [14] has broadly received attention. Because anomalous sounds of electronic motors can be used to identify conditions of fault. Thus, ASD early detection can prevent problems of electronic motors. ASD is the task of distinguishing between normal and anomaly states based on the sounds produced by target mechanical components [15], [16]. This method aims to detect anomalies early on to prevent mechanical failures. However, in real-world scenarios, extremely rare cases exist wherein machines produce anomaly sound data, leading to incorrect fault diagnosis. Therefore, identifying anomaly patterns in all cases is difficult.

To overcome these limitations, unsupervised ASD research has attracted considerable attention [17], [18], [19]. Unsupervised ASD exhibits considerable advantages in datasets without labeled data and can be used to differentiate between normal and anomaly sound patterns [20], [11], [21]. Typically, in unsupervised methods, the majority of the dataset is assumed to represent normal patterns, and anomaly patterns can be identified by comparison with normal patterns [22]. Furthermore, in unsupervised learning approaches, normal data are first learned and various metrics are subsequently used to identify anomaly data.

These methods can be categorized into three main types: distribution-based methods, distance-based methods, and reconstruction-based methods [23], [22]. Distribution-based methods rely on the statistical properties of the data to model the normal behavior of the system. Distribution-based methods assume that normal data points follow a certain statistical distribution. Anomalies are subsequently identified as data points that deviate significantly from this modeled distribution. Gaussian Mixture Models (GMMs) [24], [25] and Kernel Density Estimation (KDE) [26] are examples of distribution-based approaches. In distance-based methods, anomalies are identified based on the notion of similarity or dissimilarity (distance) between data points. In these methods, data points that are significantly distant from their nearest neighbors or from the majority of the data points are considered anomalies. K-Nearest Neighbor (KNN) [27] and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [28] are prominent examples of distance-based anomaly detection techniques. Reconstruction-based methods, typically associated with deep learning, involve training a model to reconstruct input data effectively. After training, the model can reconstruct normal data accurately but can not do that for anomalous data. The assumption is that the model learns the pattern of normal data during training, and deviations from these patterns result in higher reconstruction

errors. Autoencoders (AE) [29], [30], [31], [32], [33], [34], [35] are a prime example of this approach.

However, the field of ASD research on EPS electronic motors continues to face limited availability of labeled anomaly data on EPS electronic motors, which is a major obstacle to ASD research. Hence, it is challenging to predict all possible anomalous in each conditions and to define potential defects range. This limitation hinders the training and validation of predictive models. As automotive safety demands continue to evolve, it is important to overcome the limitations of automotive anomaly detection.

This study highlights a novel approach for ASD in EPS electronic motors. This method is an ASD method that uses acoustic data measuring defects in the outer and inner rings of bearings for EPS electronic motors. To overcome the limitations of anomaly detection, this study proposed an acoustic feature extraction method that can effectively highlight the feature differences between normal and anomaly sound data generated from an EPS electronic motor. The proposed method is focused on the unique characteristics of EPS motor sounds and prioritizes feature extraction and visualization for early anomaly detection using an unsupervised Long Short-Term Memory Autoencoder (LSTM-AE) approach. The feature extraction method has two techniques. First, the Mel Frequency Cepstral Coefficients (MFCC) [36] extraction method is used to concisely represent the sound power spectrum to eliminate irrelevant noise and focus on the most crucial aspects. In this technique, the strengths of human-like sound frequency perception and efficient encapsulation of sound characteristics is used to extract valuable information necessary for anomaly sound detection. Subsequently, the proposed dynamic feature maximization transformation (DFMT) extraction method amplifies the difference between normal and anomaly sounds. This strategy focuses on the raw amplitudes and waveform information of acoustic signals, supplementing certain acoustic details possibly overlooked by MFCC, which ensure the unique acoustic characteristics. The combination of MFCC and DFMT assigns higher scores to anomaly acoustic data, facilitating discrimination of unclear data during the model training process. We proposed an anomaly detection method using LSTM-AE based on the reconstruction error rate by extracted features. In the AE, it operates by reconstructing data instances and learning their representation in a latent space, aiming to amplify the distinction between normal and anomaly samples. As a model proficient in processing sequential data, LSTM is useful for analyzing time-series data in scenarios wherein understanding temporal dynamics is crucial for accurate anomaly detection. Therefore, we use LSTM-AE to inherit the reconstruction of the AE capabilities when using the ability of LSTM to process sequential data. Additionally, the latent vectors of the latent space of LSTM-AE are distributed to a low-dimensional space. Thus, this method can clearly evaluate the degree of anomaly by visualizing the distribution between normal and anomaly samples in a low-dimensional

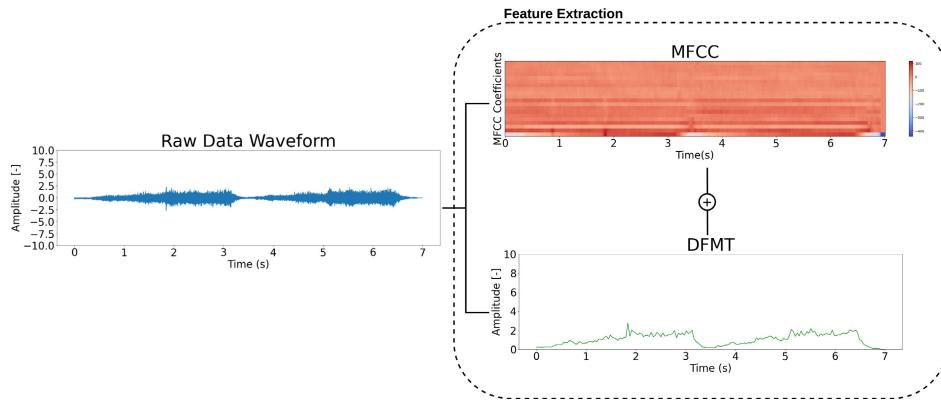


FIGURE 1. Feature extraction example from the raw signal sound to MFCC and DFMT.

space. The latent space functions as a bottleneck in the network and forces the LSTM-AE to learn the most important features of the data. The concept of using latent vectors for visualization is not entirely new, but we show that using AE architectures such as integration with LSTM for sequential data, we can intuitively identify the degree of anomaly that is lacking in traditional anomaly detection. Finally, the methodology proposed in this study provides high accuracy in early diagnosis and detection of anomaly conditions.

The contributions of this study are as follows:

- MFCC and DFMT Extraction Proposal: We introduced the MFCC and DFMT extraction methods for distinguishing between normal and anomaly motor acoustic data. This approach enhances feature representation and can capture critical sound characteristics essential for accurate anomaly detection.
- Application of Features to LSTM-AE: By applying the extracted features of MFCC and DFMT to LSTM-AE, we detect anomalies based on reconstruction error rates. This integration leverages deep learning to refine anomaly detection accuracy through temporal data analysis.
- Visualization Using Latent Vectors of the Latent Space of AE: Using the latent vectors of the LSTM-AE latent space, we provide visualizations that illustrate the distribution and degree of anomalies compared with normal data. This visualization can help in understanding complex anomaly patterns and substantiates the model capabilities.

The rest of the paper is organized as follows. Section II describes the proposed method in detail. Section III presents the results of the proposed study, providing a detailed examination of this experiments. Finally, Section IV presents a conclusion of the paper, summarizing the primary contributions and suggesting directions for future research.

II. METHODOLOGY

Each step is described in Subsection A and B, which focus on the principles underlying the proposed method. Section B

provides an explanation of the LSTM-AE by using the proposed feature extraction technique. Figure 4 illustrates the pipeline of the proposed method.

A. FEATURE EXTRACTION OF THE ELECTRONIC MOTOR SOUND SIGNAL

Feature extraction from sound signals is crucial in audio analysis for the identification and isolation of relevant characteristics from complex audio data [37], [38]. This process reduces the audio data dimensionality, making it more manageable for algorithms to process and analyze. By extracting features such as frequency, amplitude, and temporal properties, researchers can effectively characterize and distinguish between different types of sounds or audio signals. In this study, two sound features were used, including MFCC and DFMT. Figure 1 shows the feature extraction of the proposed method.

1) MFCC IMPLEMENTATION

We focused on extracting the electronic motor sound signal using MFCC. MFCC is a well-established function widely used in various speech analyses such as speech recognition and emotion analysis. However, MFCC can also be used for anomaly detection and diagnosis [39], [40], [41], [42], [43]. MFCC contains both time and frequency information of the signal, making it extremely useful for feature extraction. Additionally, MFCC can handle dynamic features by extracting both linear and nonlinear properties of the signal. Because motor sound signals contain both linear and nonlinear features, MFCC can be an effective tool to extract features of motor rotation acoustic signals. MFCC depicts the short-time power spectrum of an audio clip through the discrete cosine transform of the logarithmic power spectrum, which is mapped onto a nonlinear Mel scale. MFCC demonstrates varying sensitivity across various frequency ranges, with an acute response in lower frequencies. By analyzing these coefficients, systems can detect anomalies with higher accuracies, which enable timely maintenance and prevent

potential machinery breakdowns, ensuring efficiency and safety.

The process to compute MFCC is methodical and involves the following stages: framing the signal, calculating the power spectrum, applying Mel filter banks, computing the logarithmic values of filter banks, and applying the Discrete Cosine Transform (DCT). Figure 2 depicts the mathematical procedure of the MFCC. The core of the feature extraction process centers around the extraction of 20 MFCCs, a widely used representation for characterizing the spectral content of voice signals [44]. The details of MFCC feature extraction are as follows:

1) Pre-emphasis:

Pre-emphasis is used to compensate for the reduction in energy of high-frequency components. As a first step in MFCC, this can be adopted by simply applying a high-pass filter, typically with settings of pre-emphasis coefficient values α between 0.95 and -1. The pre-emphasis is expressed by following equation:

$$s'_n = s_n - \alpha \cdot s_{n-1}, \quad (1)$$

where the dataset sample is $s_n, \{s_n, n = 1, \dots, N\}$ [45] in each window, and the value of α is 0.97 in our case.

2) Framing:

Framing is performed to form vectors, which are the fundamental units of analysis in the speech recognition process, typically 20–30 ms long samples set into one frame.

3) Windowing:

Windowing is a crucial technique in signal processing, particularly sound signal analysis. In this method, a continuous speech signal is divided into short, fixed-length segments called frames. The purpose of applying windowing functions is to smooth out discontinuities at frame boundaries. Among various window functions, the Hamming window is widely used in connection with MFCC extraction. A smooth transition to zero occurs at both ends, effectively maintaining continuity between the first and last points within each frame. A window function $w(n)$ is expressed by the following equation:

$$s'_n = \left\{ 0.54 - 0.46 \cos \left(\frac{2\pi n}{N-1} \right) \right\} \cdot s_n, \quad (2)$$

where $0 \leq n \leq N-1$ [45].

4) Fast Fourier Transform:

After applying the Hamming window, the signal undergoes Fourier transform to transform from the time series to the frequency domain, changing the representation while retaining all relevant information. In this process, the time domain signal is converted to a frequency domain signal.

5) Mel filterbank:

Nonlinear Mel scale is applied to the converted data using a triangular filter bank. The Mel is calculated by

following equation:

$$f_{\text{mel}} = 2595 \log_{10} \left(1 + \frac{f}{700} \right), \quad (3)$$

where f is the actual frequency in Hz.

6) DCT:

In this step, the logarithm of the Mel filterbank energies $\log(f_{\text{mel}})$ is computed. This transformation to the logarithmic scale accurately reflects the human perception of sound intensity. Subsequently, DCT is applied to transform the log Mel spectrum back into the time domain. This operation creates the MFCC, which is a collection of coefficients known as acoustic vectors. Consequently, each sound input is represented by a set of acoustic vectors, providing a basis for analyzing sound patterns. The DCT is defined by the following equation:

$$c(l) = \sqrt{\frac{2}{N}} \sum_{d=1}^N \Omega_d \cos \left[\frac{\pi l}{N} \left(d - \frac{1}{2} \right) \right], \quad (4)$$

where N is the number of filter bank channels. MFCC is calculated from the log filterbank amplitudes $\{\Omega_d\}$ using the DCT [45].

This process effectively transforms the log Mel spectrum into a set of MFCC $c(l)$ which is a concise and useful function for sound signal characterization. The results of the MFCC dataset of this study comprise 20 frames, each containing 351 features. Subsequently, a standard scaler is applied to the MFCC to ensure uniformity across the feature set. Using a standard scalar is crucial in scenarios involving algorithms sensitive to the magnitude of input features, particularly when the features approximate a standard normal distribution.

2) DFMT

In this study, instead of various existing envelope detection methods, we propose a new method that can effectively track amplitude fluctuations while reducing the dimensionality of the data. This technology enhances ASD accuracy by combining with MFCC. The DFMT method is used to compensate for certain missing acoustic details in MFCC. The proposed method preserves all features and reflects the amplitude without removing important noise from the data. In particular, it preserves both low- and high-frequency characteristics by selecting the maximum amplitude within a specific window size. Due to this ability to effectively enhance feature representation by maximizing relevant features while minimizing noise, we refer to the method as a DFMT.

The main advantages of the proposed method are as follows. First, it preserves the features of the data as much as possible. To achieve this, we divided the data into specific window sizes and selected the maximum amplitude of each window size to reflect the overall amplitude change in the data. It also effectively preserves frequency characteristics. Rather than simply selecting the high amplitude, both

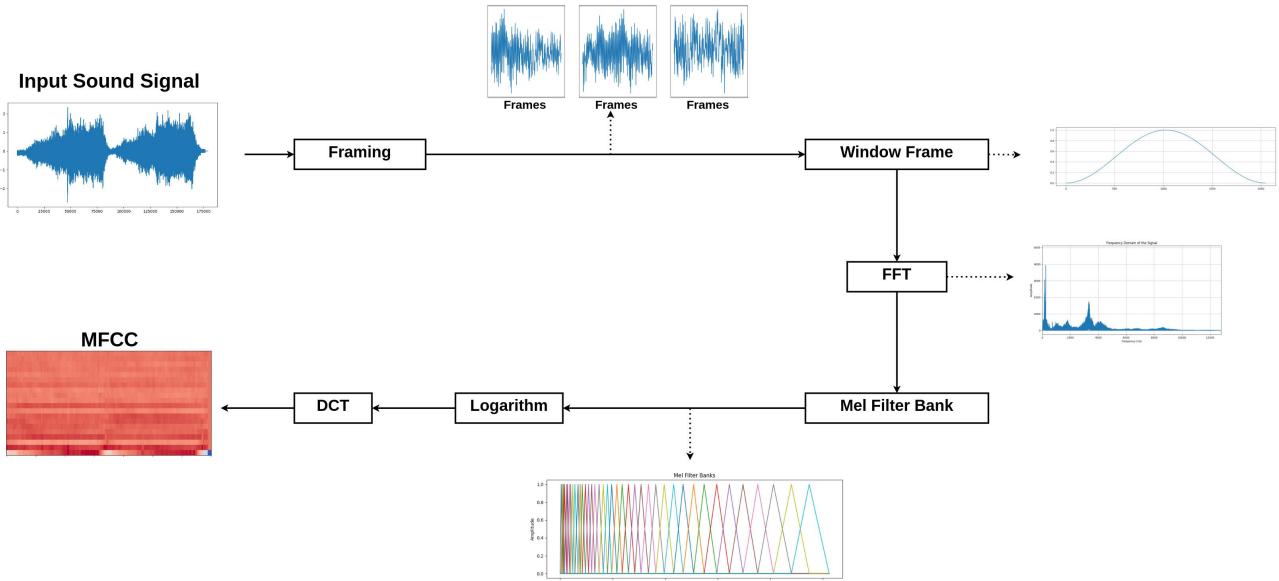


FIGURE 2. MFCC process. First, framing the signal. Next, calculating the power spectrum, applying the Mel filter bank, computing the logarithmic values of the filter bank, and finally applying the DCT.

low- and high-frequency characteristics can be maintained by comparing each section and extracting the maximum amplitude. Additionally, the proposed method provides a dimensionality reduction function. In contrast to the Hilbert transform, this method can increase processing efficiency by tracking amplitude fluctuations while reducing the dimensionality of the data. Finally, it has low noise sensitivity. The Hilbert transform is sensitive to high-frequency noise, which can amplify even small amplitude fluctuations, but the proposed method is immune to this effect, solving the inaccurate envelope problem.

In the first step of the DFMT calculation, all signals are rectified to positive values. Rectifying all signal values ensures that they are positioned within the positive domain, which enhances the clarity of observing signal variability. By eliminating negative values, the feature extraction process for measuring data variability is simplified. Additionally, this approach can be used to detect subtle changes that can occur in anomalous states. For instance, it can be used when calculating statistical properties such as the mean and variance of amplitude, ensuring all values are positive enables consistent computations. Moreover, DFMT provides a method to reduce the number of features and maintain the waveform of the original intensity. The waveform information of acoustic intensity encompasses the fundamental characteristics of sound. Preserving this information considerably improves anomaly detection accuracy by maintaining the sound variations or anomalies. Additionally, waveform information enables the tracking of temporal changes in sound, which is crucial for intuitively representing changes in the motor condition or performance degradation over time. Furthermore, anomaly detection typically exhibits characteristics that are pronounced at high amplitudes.

Preserving the waveform information of the acoustic intensity and retaining values with high amplitudes that are anticipated to be anomalous with the elimination of unnecessary sounds of low intensity simplifies data interpretation. This strategy mitigates the risk of overfitting, which can occur when the model relies excessively on a large number of features.

Let the sound signal dataset $X = \{x^{(\text{norm})}, x^{(\text{anom})}\}$, where $x^{(\text{norm})}, x^{(\text{anom})} \in \mathbb{R}^{N \times m}$, represent the normal (norm) and anomalous (anom) samples, respectively, with N denoting the number of samples and m denoting the number of features.

$$f_{\text{abs}} = |X| \quad (5)$$

Next, we defined a window size, ws , as 1,000. Let X^* be a new dataset constructed as follows:

$$X^* = \left\{ \max_{j \in [(k-1) \cdot ws + 1, k \cdot ws]} f_{\text{abs}} x_i^{(j)} : k = 1, 2, \dots, \lceil \frac{m}{ws} \rceil \right\}, \quad (6)$$

for input data $x_i^{(j)}$ in subset of X , where $j = 1$ to m and $i = 1$ to N . This process result in the preparation a new dataset of features X^* . This formulation ensures that X^* consists of data samples with reduced dimensionality, where each data sample in X^* retains the maximum values from each ws , effectively simplifying the dataset and preserving crucial information for anomaly detection. Figure 3 shows the process of the DFMT. Figure 3 (a) shows the original signal sample of the sound data. Figure 3 (b) depicts the range of negative and positive amplitude values. In Figure 3 (c), the negative amplitude value is converted to a positive amplitude value. Figure 3 (d) depicts the maximum value in each section by dots. Figure 3 (e) shows the graph with the final maximum point value applied. Compared with the waveform of the positive amplitude values of the original sound signal, the

values drawn with DFMT demonstrate that important features of the data can be revealed when preserving the waveform shape as much as possible.

The mathematical formula for defining a new dataset combining DFMT features and MFCC features is as follows:

To align the MFCC features with DFMT, Let X^* denote a dataset obtained by extracting DFMT features, and let C^* denote a dataset obtained by applying a standard scalar to the MFCC features extracted from the original data. The new dataset Y is constructed by concatenating X^* and C^* , represented as follows:

$$Y = \text{concat}(X^*, C^*) \quad (7)$$

This dataset Y is used as the input for the LSTM-AE.

B. ANOMALY DETECTION USING LSTM-AE

In this section, we introduce the proposed model combining LSTM and AE to detect anomalies based on sound data analysis. Figures 5 and 6 depict the AE and LSTM structures. Figure 5 illustrates a typical LSTM unit, comprised of cells, input gates, output gates, and forget gates. The LSTM-AE is designed for anomaly sound detection by learning to reconstruct normal sound sequences. The model comprises two components: an encoder and a decoder, both using LSTM layers, followed by a linear layer to ensure that the output shape matches the input shape. The detail of LSTM is explained as follows:

1) LSTM

1) Forget Gate:

The forget gate is used to evaluate what information in the cell state should be kept or discarded based on the current input and previous hidden state. The sigmoid function is used to create a vector with values between 0 and 1 that represent the importance of each piece of information. A value close to 0 indicates that the information is not relevant and should be forgotten, while a value close to 1 indicates the information is relevant and should be retained. The equation for the output f_t of the forget gate is as follows:

$$f_t = \sigma(w_f \cdot [H_{t-1}, X_t] + b_f), \quad (8)$$

where σ denotes the sigmoid activation, w_f and b_f are the weight and bias of the forget gate, respectively, and H_{t-1} and X_t are the previous hidden state and current input respectively.

2) Input Gate:

The input gate has twofold purposes. First, the input gate evaluates the relevance of the new inputs for updating the cell state. Second, the gate determines the components of new data to be stored. This function is performed in two steps, summarized in (9) and (10). In the first step, the memory update vector \tilde{C}_t is calculated from the combination of the previous hidden state and the incoming data and adjusted

by a hyperbolic tangent function to produce values between $[-1, 1]$. This vector quantifies the extent of modification each cell state component should undergo for new data, as presented in the following equation:

$$\tilde{C}_t = \tanh(w_c \cdot [H_{t-1}, X_t] + b_c), \quad (9)$$

$$i_t = \sigma(w_i \cdot [H_{t-1}, X_t] + b_i), \quad (10)$$

where in (9) w_c and b_c are the weight matrices and the bias of the input gate, respectively, and a tanh is used as an activation function. In (10) w_i and b_i are the weight matrices and the bias of the input gate, respectively. The pointwise multiplication is then processed to update the cell state.

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (11)$$

3) Output Gate:

The primary function of the output gate is to establish the new hidden state by using three key pieces of information: the updated cell state, the previous hidden state, and the new input data. This process involves passing the previous hidden state and current input data through a sigmoid-activated network to generate the filter vector σ , as outlined in (12). This step is crucial for determining which parts of the cell state will be used to form the new hidden state, streamlining the LSTM process of updating its long-term memory.

$$o_t = \sigma(w_o \cdot [H_{t-1}, X_t] + b_o), \quad (12)$$

where w_o and b_o are the weight matrix and the bias of output gate, respectively.

The LSTM unit processes input data sequentially, where each unit updates the cell state C_t and the hidden state H_t . On completion, C_t and H_t are converted to C_{t-1} and H_{t-1} for the subsequent unit, facilitating a continuous flow of information through the network. This iterative process ensures comprehensive data processing across the time series. The output gate modulates the cell state using the filter vector o_t , post-activation, to generate the new hidden state. This activation function confines the values within the $[-1, 1]$ range, with in (13) detailing the new hidden state formulation. Consequently, each new state feeds into the next LSTM unit, maintaining the network memory and predictive capabilities throughout the sequence.

$$H_t = o_t \odot \tanh(C_t) \quad (13)$$

2) AUTOENCODER(AE)

AE is an unsupervised neural network used to learn efficient encoding of unlabeled data. AE is trained to ignore irrelevant data to learn a representation for the data set. Structurally, AE consists of an input layer, an output layer, and several hidden layers. AE operation can be categorized into encoding, decoding, and reconstruction loss calculation, as shown in Figure 6.

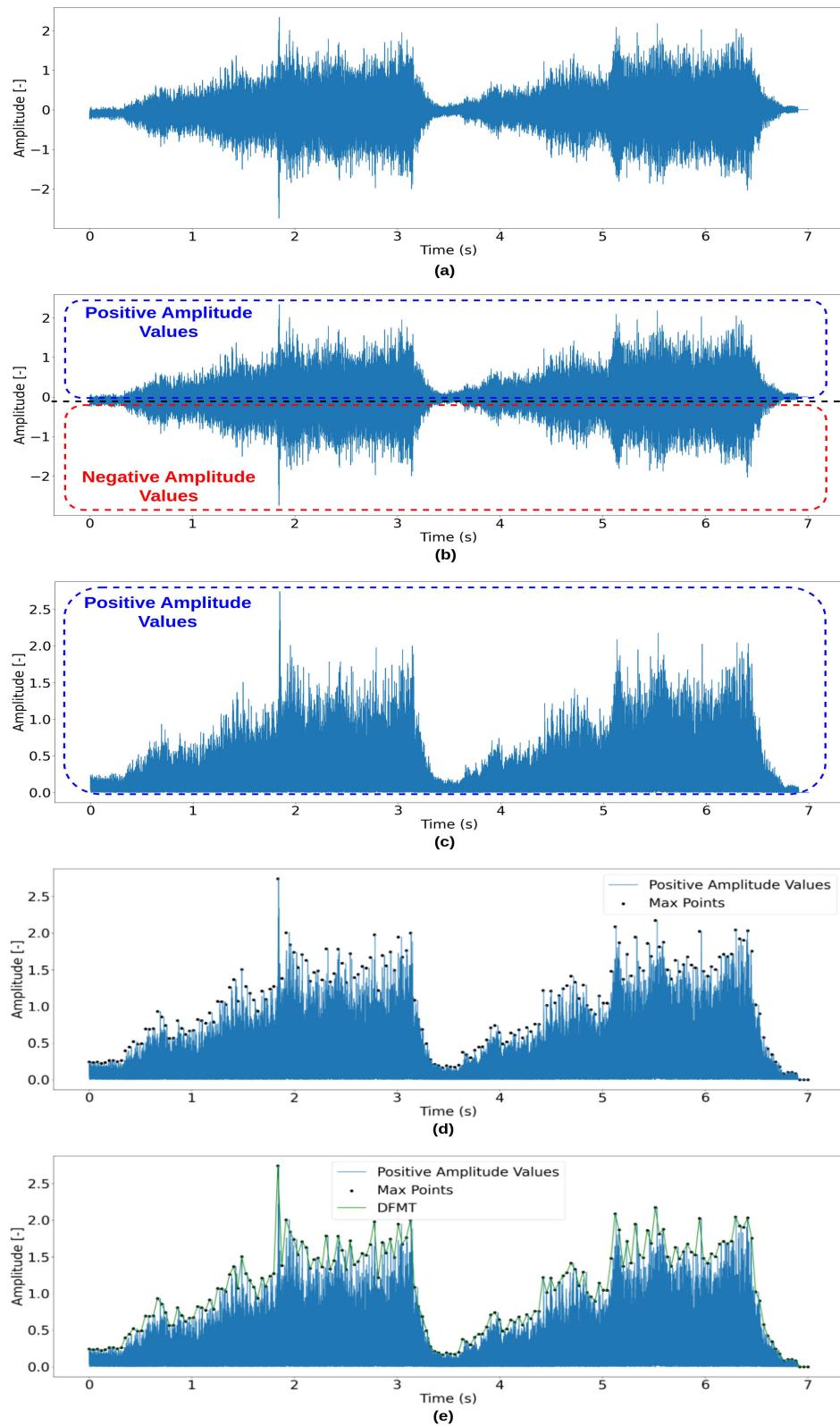


FIGURE 3. Process of the DFMT. Figure 3 (a) displays the original sound signal of the data. (b) depicts the range of negative and positive amplitude values. In (c), the negative amplitude value is converted to a positive amplitude value. (d) shows the maximum value in each section by dots. (e) finally shows the graph with the final maximum point value applied. Compared to the waveform of positive amplitude values of the original signal, the values drawn with DFMT demonstrate that important features of the data can be revealed while preserving the waveform shape as much as possible.

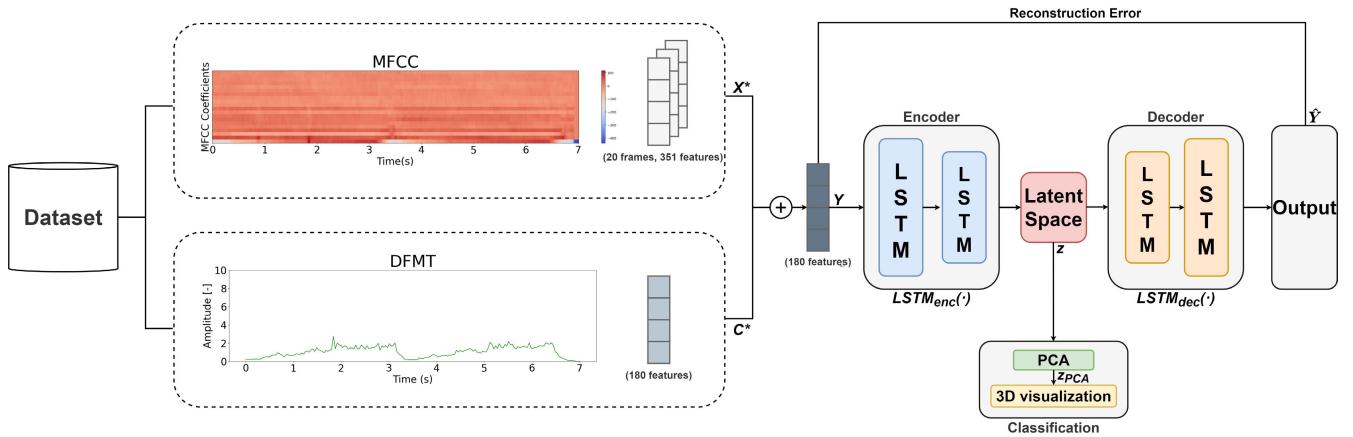


FIGURE 4. Pipeline of the proposed MFCC and DFMT LSTM-AE. The figure shows the process starting with the raw data in the dataset, performing feature extraction to generate MFCC and DFMT features. These features are then merged into a unified dataset Y , which contains selected features compatible with the LSTM-AE input requirements. LSTM-AE consists of an encoder that reduces data dimensionality to a latent space Z and a decoder that reconstructs the signal and calculates the reconstruction error. Latent spaces can be used for dimensionality reduction through PCA to facilitate 3D visualization and classification.

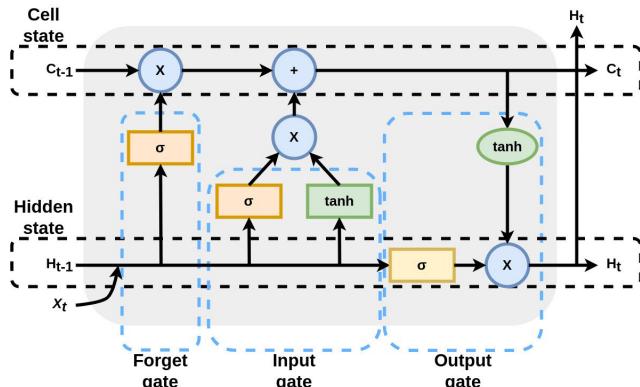


FIGURE 5. The architecture of LSTM.

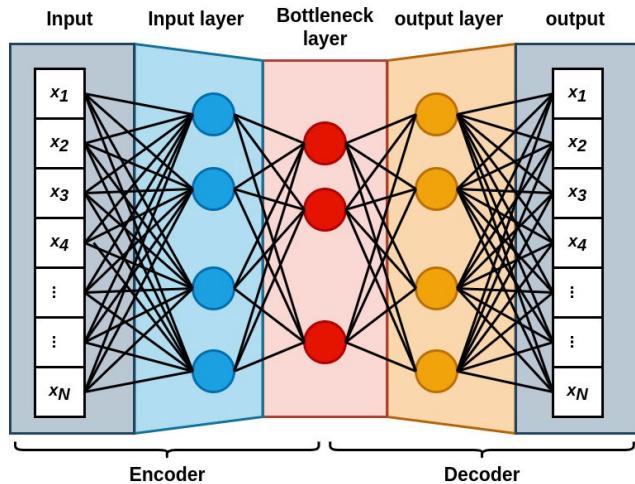


FIGURE 6. The architecture of AE.

1) Encoding:

The encoding operation maps input data y , which is a high-dimensional vector $y \in \mathbb{R}^t$, to a low-dimensional

bottleneck layer representation (h).

$$h = f_1(w_i y + b_i), \quad (14)$$

where w_i is the weight matrix, b_i is a bias, and f_1 is an activation function.

2) Decoding:

In this case, the bottleneck layer representation of h generates the output \hat{y} , which attempts to reconstruct y .

$$\hat{y} = f_2(w_j h + b_j), \quad (15)$$

where f_2 is an activation function for the decoder, w_j is the weight matrix, b_j represents a bias, and \hat{y} represents the reconstructed input sample.

3) Reconstruction Loss:

In (16) illustrates how reconstruction loss (L) is calculated in a typical AE to minimize the difference between the output and the input.

$$L(y - \hat{y}) = \frac{1}{n} \sum_{t=1}^n |\hat{y}_t - y_t|, \quad (16)$$

where y represents input data, \hat{y} indicates output data, and n is the number of samples in the training dataset.

The LSTM-AE is trained on sequences of normal sounds to minimize the reconstruction error, thereby learning a compact representation of normal sound patterns. During the inference phase, sequences that exhibit a high reconstruction error are flagged as anomalies, indicating potential sound anomalies. Therefore, if the input data are significantly different from the training data (i.e., anomalies), AE will not be able to reconstruct it accurately, leading to high reconstruction errors. By monitoring the magnitude of these reconstruction errors, anomalies can be detected. Data points with high reconstruction errors are considered anomalies. The trained

AE model is expected to reconstruct generic input data with extremely low reconstruction errors. However, it does not function well with anomalous data because it has not learned how to handle these deviations. Therefore, training the model using only normal data is important [46]. If we use anomalous data in our training model, training with anomalous data may cause the model to specialize in detecting only the types of anomalies present in the training set, reducing its ability to detect new or unexpected anomalies. By using only normal data for training, the model develops a more generalized understanding of normal patterns, improving its ability to identify deviations from these patterns as anomalies. This method facilitates effective anomaly detection in sound data by using the temporal dynamics that the LSTM layers learn.

The use of LSTM layers enables the model to capture long-term dependencies within the sound sequences, which is essential for discerning complex patterns indicative of normal and anomalous states. The encoder-decoder architecture aids in learning a condensed knowledge of the sound data, enabling the detection of deviations from the learned normal pattern.

C. VISUALIZATION

To visualize anomaly detection, we applied the Principal Component Analysis (PCA) [47] in the latent vectors of the LSTM-AE latent space to render easy identification of normal and anomaly patterns. PCA is a widely recognized and effective technique used in compression and noise reduction [48]. PCA transforms high-dimensional data into a low-dimensional subspace with retention of as much variance as possible. This phenomenon renders easy visualization and interpretation of relationships between data samples, including distinguishing between normal and anomaly patterns. PCA filters noise and less useful features from the latent space by focusing on the principal components that capture the largest changes. Visualizing principal components can provide insight into the underlying structure of the data, revealing outliers or patterns that can not be identified in a high-dimensional space. When visualizing latent vectors in three dimensions, the visualization can be under-informative if it does not include important aspects of the data needed to distinguish anomalies. Therefore, by applying PCA to a larger dimensional latent space, we can reduce the dimensionality and extract the most important patterns while considering the features and relationships of a wide array of data. This methodology not only uncovers complex patterns within the dataset but also effectively highlights the differences between normal and anomalous data features, with PCA playing a crucial role in identifying features indicative of anomalies [49]. Therefore, the latent vectors of LSTM-AE were considered and PCA was applied to visualize data samples for anomaly detection.

1) Definition of Latent Vectors:

Let h_i represent the latent vector obtained from the encoder for the i -th data sample, defined

as follows:

$$h_i = f_1(w_i y_i + b_i), \quad (17)$$

where f_1 is the activation function of the encoder, w_i represents the weight matrix of the encoder, b_i denotes the bias term, and y_i is the input data sample.

2) Dimension Reduction using PCA:

a) Mean Centering:

Calculate the mean of the latent vectors:

$$\bar{h} = \frac{1}{n} \sum_{i=1}^n h_i \quad (18)$$

Center the latent vectors by subtracting the mean:

$$h_{i,\text{centered}} = h_i - \bar{h} \quad (19)$$

b) Covariance Matrix Computation:

Compute covariance matrix C from the centered latent vectors:

$$C = \frac{1}{n-1} H_{\text{centered}}^T H_{\text{centered}}, \quad (20)$$

where H_{centered} is the matrix of all centered latent vectors.

c) Eigenvalue Decomposition:

Perform eigenvalue decomposition on the covariance matrix C to obtain eigenvectors V and eigenvalues Λ :

$$CV = V\Lambda \quad (21)$$

d) Select Principal Components and Project Data:

Select the top k eigenvectors to form a projection matrix V_k . Project the centered latent vectors onto the lower-dimensional space to obtain new latent vectors $h_{i,\text{PCA}}$:

$$h_{i,\text{PCA}} = h_{i,\text{centered}} V_k \quad (22)$$

After PCA application, the dataset in the reduced latent space, $h_{i,\text{PCA}}$, contains samples existing in a three-dimensional space, suitable for visualization and analysis.

III. EXPERIMENTS AND RESULTS

A. DATASET

In this study, data obtained from measuring the sound of electronic motors among EPS in the automobile industry were used for the experiments. This dataset was measured at the quality control line of an industrial manufacturing process, and the conditions were the same as experimental measurement equipment. The electronic motor dataset was collected for both clockwise (CW) and counterclockwise (CCW) rotations of the motor, with each direction recorded for a duration of 3.5 s. Consequently, a total of 7 s of sound data were measured per motor. Additionally, each sound data was represented as a vector with 179,200 features. The dataset

consists of 29,759 data sets, comprising 29,294 normal and 465 anomaly data. The causes of defects are varied and can be classified into several types. The cause of the defect can be a CW or CCW problem, and the degree of CW and CCW defects can differ. Furthermore, the dataset includes data indicating conditions that are difficult to determine regardless of whether it is normal or anomaly and is used to establish standards for improving the automation and speed of defect judgment. The dataset was partitioned into training and testing sets with an 80:20 ratio. The training dataset consists exclusively of normal data samples, encompassing 23,435 samples. By contrast, the testing dataset comprises 6,324 samples with anomaly data samples. This division ensures a comprehensive evaluation framework, allowing for the assessment of the ability of the LSTM-AE to generalize from the learned representations of normal patterns to identify anomalies in unseen data.

TABLE 1. Mel & MFCC & DFMT parameters.

Parameters	Values
n_mel	128
n_fft	2048
hop_length	512
n_mfcc	20
sr	25.6 kHz
win_size	1,000

B. EXPERIMENTAL SETUP

In the experimental setup of the study, we meticulously selected parameters for feature extraction to ensure the optimal performance of the proposed LSTM-AE classifier. The feature extraction process plays a crucial role in the efficacy of the classifier, particularly when handling audio signals. In the experiments, we focused on extracting features using Mel Spectrograms, MFCC, and DFMT. Table 1 lists the parameters for each of these feature extraction methods.

We set the number of Mel bands to 128, which dictates the resolution of the Mel scale. This parameter influences the Mel spectrogram by determining how the frequency spectrum is divided into Mel bands. The Fast Fourier Transform (FFT) size was set to 2,048, defining the length of the windowed signal after padding with zeros. A larger n_fft increases the frequency resolution of the resulting spectrogram. The hop length, which is the stride with which the window is moved across the signal, was set to 512 samples. This parameter affects both the time resolution and the overlap between successive frames in the spectrogram. We extracted 20 MFCC features from each audio frame. MFCCs are widely used in audio processing because they effectively represent the power spectrum of sound. The audio signals were processed at a sampling rate of 25.6 kHz. This low sampling rate was used to reduce computational complexity while retaining sufficient information for feature extraction. The window size for DFMT was set to 1,000. This window size is critical for capturing the temporal characteristics of the audio signal

within each frame. By optimizing these feature extraction settings, we captured the essential characteristics of the audio signals, which are pivotal for accurate anomaly detection using the proposed LSTM-AE classifier.

C. PERFORMANCE METRICS

To evaluate the performance of the proposed method, we used several metrics: accuracy, precision, recall, and F1 score. These metrics provided a comprehensive evaluation of performance.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (23)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (24)$$

$$\text{Recall}(= TPR) = \frac{TP}{TP + FN} \quad (25)$$

$$F1 - \text{score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (26)$$

$$FPR = \frac{FP}{TP + TN + FP + FN} \quad (27)$$

- **True Positives (TP):** The number of positive dataset correctly classified as positive.
- **True Negatives (TN):** The number of negative dataset correctly classified as negative.
- **False Positives (FP):** The number of negative dataset incorrectly classified as positive.
- **False Negatives (FN):** The number of positive dataset incorrectly classified as negative.
- **True Positive Rate (TPR),** also known as *Recall* or *Sensitivity*, is defined as:

$$TPR = \frac{TP}{TP + FN} \quad (28)$$

- **False Positive Rate (FPR)** is defined as:

$$FPR = \frac{FP}{FP + TN} \quad (29)$$

- **The Receiver Operating Characteristics (ROC) curve** is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. The curve is created by plotting the TPR (Recall) against the FPR for various threshold settings.
- **The Area Under the Curve (AUC)** of the ROC plot is a measure of the ability of the classifier to distinguish between classes. An AUC of 1 indicates a perfect classifier, whereas an AUC of 0.5 indicates a classifier that does not outperform random chance.

The AUC–ROC can be mathematically represented as the integral of the TPR (Recall) with respect to the FPR over the interval from 0 to 1:

$$\text{AUC}_{\text{ROC}} = \int_0^1 TPR(FPR) d(FPR) \quad (30)$$

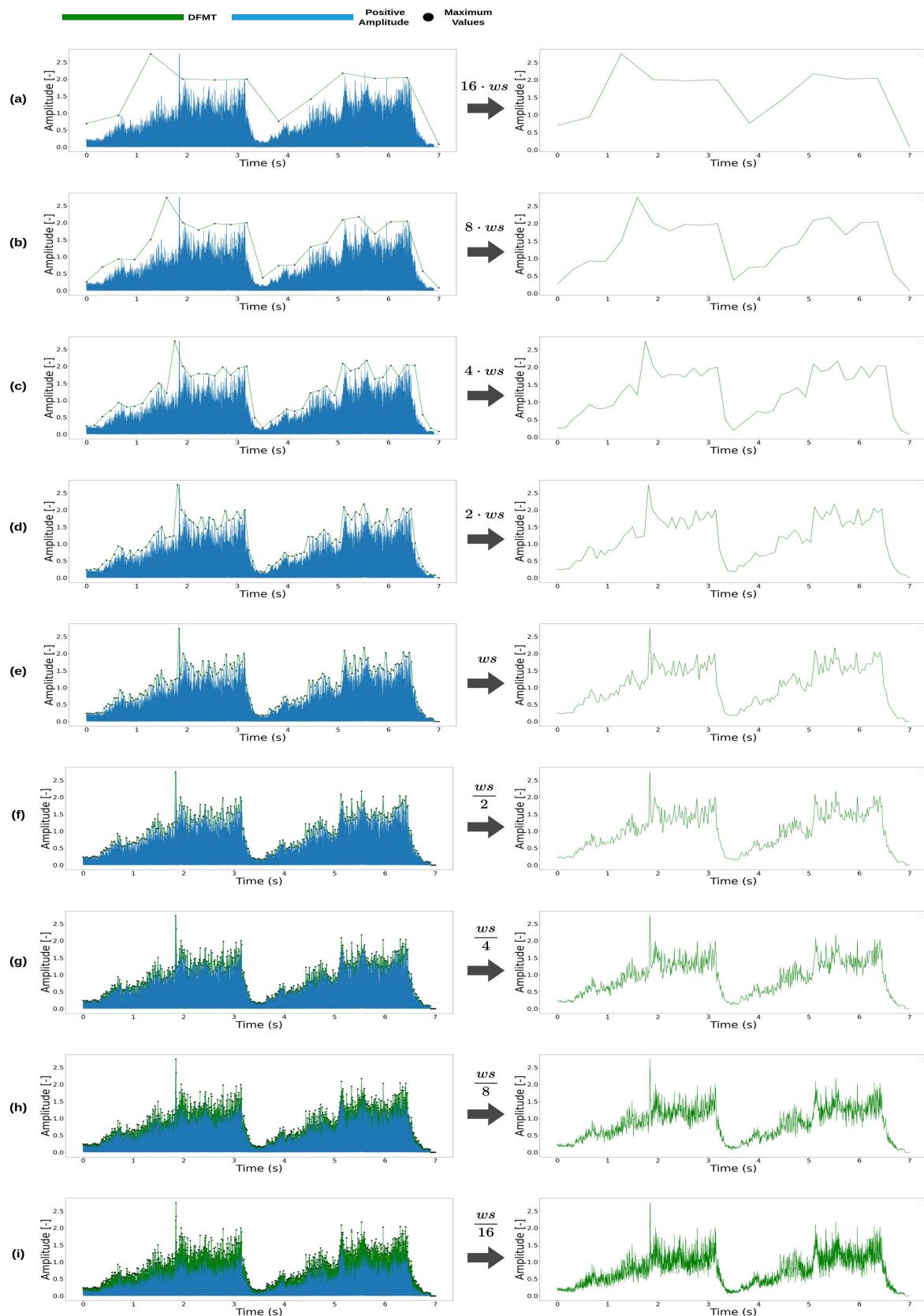


FIGURE 7. Visual representation of DFMT feature extraction across varying window sizes. (a)-(d) shows gradual increase of the window size size in DFMT. (e) DFMT Feature extraction results using the standard window size ws . (f)-(i) shows gradual reduction of the window size size in DFMT.

D. RESULTS OF THE LSTM-AE USING MFCC AND DFMT

1) COMPARISON WITH OTHER FEATURE EXTRACTIONS

To demonstrate the effectiveness of the proposed DFMT method, we provided evidence that DFMT retains the inherent shape of the data while maintaining an optimal number of features within defined window sizes. Subsequently, we compared the combination of MFCC and DFMT with other widely used method in ASD. Figures 7, 8, and Table 3 illustrate these comparisons.

Figure 7 (e) depicts the results of applying the proposed window size to DFMT, illustrating the capability of the proposed method to capture essential features. This ws plays a critical role in determining the granularity of feature extraction within the DFMT framework. Figure 7(a)–(d) shows the results of incrementally doubling the window size proposed by DFMT. Figure 7(a) shows the result of increasing ws by 16 times, and Figure 7(d) shows the result of increasing ws by 2 times. Figure 7(f) presents the outcome of doubling down the proposed window size in DFMT, examining the effect of window size adjustment on feature extraction. Subsequently, Figures 7(f)–(i) depict the results of incrementally doubling down the window size from the ws . Figure 8 shows how changing the ws affects the functionality of the DFTM representation. These figures collectively validate that DFMT effectively retains the original data characteristics while ensuring the selection of the most relevant features for anomaly detection.

Figure 8 presents a comparative analysis of several performance metrics—Accuracy, Recall, Precision, F1-Score, and AUC—evaluated under various window sizes for DFMT. The window sizes are represented using $16 \cdot ws$, $8 \cdot ws$, $4 \cdot ws$, $2 \cdot ws$, $\frac{ws}{16}$, $\frac{ws}{8}$, $\frac{ws}{4}$, $\frac{ws}{2}$, and ws on the x-axis.

The graph in Figure 8 reveals that the window size (ws) corresponds to the highest performance across all evaluated metrics. This phenomenon indicates that a general trend of performance degradation is observed as the window size increases or decreases, which is indicated by a decrease in the metric value (Accuracy, Precision, F1-Score, AUC).

Specifically, relative to ws , the accuracy (indicated by the blue line) decreases as the window size increases from $2 \cdot ws$ to $16 \cdot ws$. However, as the window size decreases from $\frac{ws}{2}$ to $\frac{ws}{16}$ relative to ws , the accuracy remains relatively high, with ws achieving the highest accuracy at 99.0%. The recall (orange line) remains at 1.0 as the window size increases from $2 \cdot ws$ to $16 \cdot ws$ relative to ws . However, relative to ws , it recalls a peak of 98.0% at $\frac{ws}{2}$ before decreasing as the window size decreases from $\frac{ws}{2}$ to $\frac{ws}{16}$. This means that as the window size increases, the rate of misclassifying anomaly data as normal decreases, but the rate of misclassifying normal data as anomaly increases. Conversely, as the window size becomes smaller, the rate at which the model misclassifies normal data as anomaly decreases, but the rate at which it misclassifies anomaly data as normal increases. Even though the recall of $\frac{ws}{2}$ is higher than ws , the following Precision and F1-Score results show that ws is the most appropriate window size. Based on ws , both Precision (green line) and F1-Score

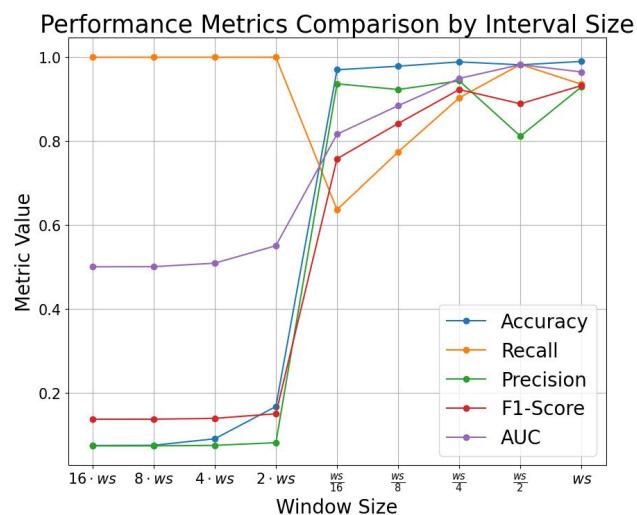


FIGURE 8. Performance evaluation of an anomaly detection across varying window sizes represented by $16 \cdot ws$, $8 \cdot ws$, $4 \cdot ws$, $2 \cdot ws$, $\frac{ws}{16}$, $\frac{ws}{8}$, $\frac{ws}{4}$, $\frac{ws}{2}$, and ws . The graph highlights the correlation between the increased window size and enhanced model performance across metrics such as accuracy, recall, precision, F1-score, and AUC.

(red line) decrease as the window size increases from $2 \cdot ws$ to $16 \cdot ws$. Similarly, reducing the window size from $\frac{ws}{2}$ to $\frac{ws}{16}$ also lowers the results, indicating that window sizes close to ws is the most suitable. Among sizes between $\frac{ws}{2}$ and $\frac{ws}{16}$, $\frac{ws}{2}$ shows the largest reduction, with a particularly noticeable drop in Precision. That is, $\frac{ws}{2}$ has a lower rate of misclassifying anomaly data as normal compared to ws , but has a much higher rate of misclassifying normal data as anomaly, making it unsuitable for anomaly detection. Therefore, as the window size increases or decreases relative to ws , the ability to provide reasonably accurate results decreases, as evidenced by a significant decrease in the measure of the trade-off between recall and F1 score. Finally, Figure 8 shows lower values relative to ws as the window size increases from $2 \cdot ws$ to $16 \cdot ws$. Similarly, reducing the window size from $\frac{ws}{2}$ to $\frac{ws}{16}$ lowers the AUC value. This indicates that the model ability to discriminate between classes decreases as the window size decreases. Although $\frac{ws}{2}$ has a higher AUC, our previous analysis of accuracy, precision, recall, and F1 score shows that ws is the best window size for anomaly detection. This is because ws provides balanced performance across all metrics, avoiding the pitfalls of misclassification that occur with different window sizes. Therefore, ws ensures a more stable and accurate anomaly detection process. Notably, the last value of the x-axis (ws) in Figure 8 is highlighted in the results, which indicates that the largest window size tested yielded the best overall performance. This key result highlights the importance of choosing an appropriate window size to maintain high model performance. Given that anomaly detection models are sensitive to the balance between false positives and false negatives, selecting a window size that ensures an optimal trade-off is crucial because overemphasizing one parameter

can lead to inadequate performance on the other parameter. The observed trend indicated that a larger window size can maintain this balance, minimizing false positives and false negatives, while effectively identifying anomalies. Thus, the performance is highly dependent on the chosen window size, with ws sizes yielding superior results.

Subsequently, the efficacy of DFMT in extracting acoustic anomaly features was demonstrated by comparing six distinct feature combinations using a LSTM-AE, as outlined in Table 3. The experimental framework maintained consistent learning rates, batch sizes, and epochs across all trials to ensure fairness in comparison. Table 2 lists the training parameters for the feature extraction parameters and LSTM-AE.

TABLE 2. LSTM-AE parameters.

Hyperparameters	Values
Input Size	1
Hidden Size	16
Sequence Length	180
Learning Rate	0.001
Batch Size	5
Epochs	100
Optimizer	Adam

The outcomes of using solely Mel and MFCC features for LSTM-AE revealed suboptimal performance in terms of recall and precision, highlighting the limitations of these features when used in isolation. By contrast, the application of the proposed DFMT under identical experimental conditions yielded considerably enhanced performance metrics, achieving Accuracy, Recall, Precision, and F1-score values of 98.9%, 93.1%, 93.1%, and 93.1%, respectively. This improvement validates the efficacy of DFMT in capturing the intrinsic waveform and focusing on high-amplitude features, which are crucial for accurate anomaly detection. However, although DFMT emphasizes waveform and amplitude, it could overlook information obtainable at subtly lower amplitude. To address this potential information loss, we demonstrated an acoustic feature extraction method that minimizes loss by combining MFCC with DFMT. This combination not only retains the benefits of DFMT but also incorporates the frequency-based insights provided by MFCC, which provides a comprehensive feature set for anomaly detection.

Remarkably, in our experiments, this method surpassed the performance of other combinations, including Mel with MFCC and Mel with DFMT. The combination of MFCC and DFMT showed superior classification capabilities, with Accuracy, Recall, Precision, and F1-score reaching 99.2%, 94.0%, 95.6%, and 94.7%.

Using MFCC and DFMT, we found the optimal configuration that provides the best performance of the LSTM-AE model to select the best combination for anomaly detection. To achieve this, we compared the performance of the models under different settings by adjusting the latent space size and the anomaly detection threshold. Table 4 presents the

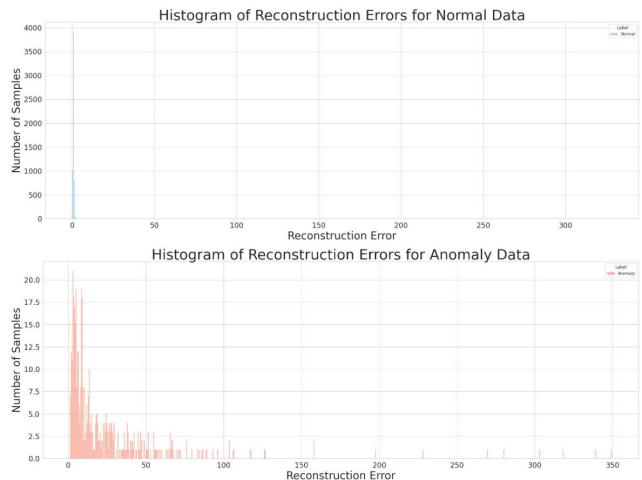


FIGURE 9. Reconstruction error of normal and anomaly data. As expected, anomaly data exhibits a large reconstruction error distribution (red).

performance comparison of the LSTM-AE models under different latent space sizes and thresholds. The performance metrics for each configuration included accuracy, recall, precision, F1-score, and AUC-ROC. This table compares the following combinations for each LSTM-AE model:

- Latent space size: 8, 16, 32, 64, 128, 256
- Threshold: 0.5, 1.0, 2.0, 3.0, 4.0, 5.0

One of the most important aspects of anomaly detection is to minimize FN (i.e., misclassifying anomaly data as normal). Therefore, models with high recall and precision are highly desirable. Among the configurations, the combination of latent space size 16 and threshold 3.0 (LSTM AE16) showed excellent performance. This particular setting achieved 99.1% accuracy, 93.5% recall, 93.8% precision, 93.6% F1-Score, and 96.5% AUC-ROC. The threshold plays a critical role in determining the performance of the model. It is important to set an appropriate threshold (in this case 3.0). If the threshold is extremely low or exceedingly high, the performance can deteriorate significantly. Our analysis confirmed that the optimal feature combination and appropriate model parameter settings are most important for maximizing the performance of the LSTM-AE model.

2) EVALUATION EXPERIMENTS

In this study, to evaluate the performance of the LSTM-AE model, we calculated the reconstruction error for each data point using the mean square error (MSE) between the original and reconstructed data:

$$\text{Reconstruction Error} = \frac{1}{n} \sum_{i=1}^n (y - \hat{y})^2 \quad (31)$$

We calculated this reconstruction error for each of the 6,324 data samples, including both normal and anomalous data. The histograms of these reconstruction errors for normal and anomalous data are shown in Figure 9 and Figure 10, respectively. The histogram of reconstruction error for normal

TABLE 3. Performance comparison of LSTM-AE classifier with various feature combinations.

Feature Combination	Learning Rate	Batch Size	Epochs	Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)	AUC-ROC (%)
Mel	0.001	5	100	89.0	51.1	33.9	40.7	71.6
MFCC	0.001	5	100	82.6	22.7	12.5	16.1	55.0
DFMT	0.001	5	100	98.9	93.1	93.1	93.1	96.2
Mel + MFCC	0.001	5	100	93.9	25.5	75.7	38.2	62.4
Mel + DFMT	0.001	5	100	95.0	41.7	82.9	55.5	70.5
MFCC + DFMT	0.001	5	100	99.2	94.0	95.6	94.7	96.8

TABLE 4. Performance comparison of LSTM-AE model with various latent space size and threshold.

LSTM AE model	Latent Space	Threshold	Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)	AUC-ROC (%)
LSTM AE8	8	0.5	17.5	100.0	8.2	15.1	55.5
LSTM AE8	8	1.0	54.8	100.0	14.0	24.5	75.6
LSTM AE8	8	2.0	98.1	99.6	79.6	88.4	98.8
LSTM AE8	8	3.0	99.0	95.3	91.0	93.1	97.3
LSTM AE8	8	4.0	98.4	85.8	91.9	88.8	92.6
LSTM AE8	8	5.0	97.7	75.5	91.6	82.8	87.5
LSTM AE16	16	0.5	17.9	100.0	8.2	15.2	55.7
LSTM AE16	16	1.0	63.9	100.0	16.9	29.0	80.5
LSTM AE16	16	2.0	98.5	98.9	83.2	90.4	98.7
LSTM AE16	16	3.0	99.1	93.5	93.8	93.6	96.5
LSTM AE16	16	4.0	98.6	85.8	94.8	90.1	92.7
LSTM AE16	16	5.0	97.9	76.3	94.4	84.4	88.0
LSTM AE32	32	0.5	14.0	100.0	7.9	14.6	53.6
LSTM AE32	32	1.0	49.7	100.0	12.8	22.6	72.9
LSTM AE32	32	2.0	96.9	99.6	70.4	82.5	98.1
LSTM AE32	32	3.0	98.7	94.4	89.0	91.6	96.7
LSTM AE32	32	4.0	98.5	88.4	90.9	89.6	93.8
LSTM AE32	32	5.0	97.8	78.5	91.0	84.3	88.9
LSTM AE64	64	0.5	7.9	100.0	7.4	13.8	50.3
LSTM AE64	64	1.0	48.1	100.0	12.4	22.1	72.0
LSTM AE64	64	2.0	96.0	99.8	65.0	78.7	97.8
LSTM AE64	64	3.0	98.8	95.7	88.5	91.9	97.4
LSTM AE64	64	4.0	98.6	90.5	90.7	90.6	94.9
LSTM AE64	64	5.0	98.2	83.9	91.3	87.4	91.6
LSTM AE128	128	0.5	7.9	100.0	7.4	13.8	50.3
LSTM AE128	128	1.0	48.9	100.0	12.6	22.4	72.4
LSTM AE128	128	2.0	96.3	99.8	66.5	79.8	97.9
LSTM AE128	128	3.0	98.8	95.7	88.6	92.0	97.4
LSTM AE128	128	4.0	98.6	90.5	90.7	90.6	94.9
LSTM AE128	128	5.0	98.2	83.4	91.3	87.2	91.4
LSTM AE256	256	0.5	7.8	100.0	7.4	13.8	50.2
LSTM AE256	256	1.0	48.6	100.0	12.5	22.3	72.3
LSTM AE256	256	2.0	96.2	99.8	66.2	79.6	97.9
LSTM AE256	256	3.0	98.7	94.6	88.4	91.4	96.8
LSTM AE256	256	4.0	98.6	90.5	90.7	90.6	94.9
LSTM AE256	256	5.0	98.2	83.0	91.3	86.9	91.2

data in Figure 9 shows that the reconstruction error for normal data is highly concentrated around zero. This distribution is extremely narrow and strongly concentrated around the specific value 0. This means that our model can accurately reconstruct normal data with minimal errors. The reconstruction error rate histogram for anomalous data in Figure 9 is distributed over a wider range. Anomaly data has various reconstruction error rates, and certain data have relatively large error values. This variance reflects the inherent diversity and complexity of anomalous data, indicating variability in the model ability to reconstruct anomalous conditions. This distribution is skewed to the right, indicating that although most reconstruction errors are small, a significant number of data samples exhibit larger reconstruction errors. This pattern represents either an exponential distribution or a Poisson

distribution. In particular, the histogram for anomaly data shows sporadic peaks at various error levels, which may indicate the presence of different types of anomalies within the data set.

Figure 10 presents a complementary view to Figure 9 by displaying a box plot of the reconstruction errors on a logarithmic scale for both normal and anomalous data. This log-scale visualization is particularly effective in differentiating the spread and central tendencies of errors between two classes. The normal data, depicted in blue, exhibit a compact box with a median close to zero, indicating that most normal data points have a low reconstruction error. The interquartile range (IQR), represented by the height of the box, is narrow, which indicates that the errors for normal data are consistently low and do not vary widely.

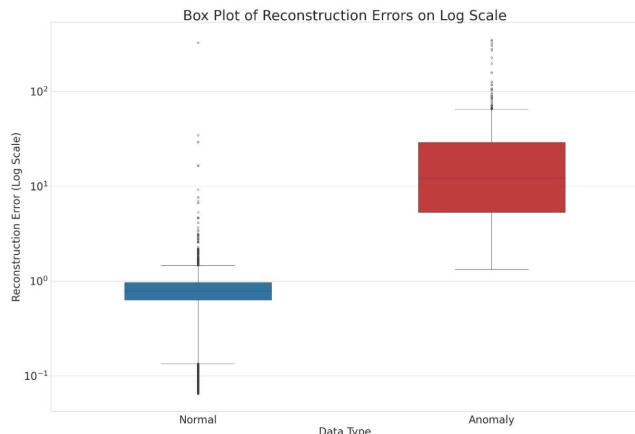


FIGURE 10. Box plot of reconstruction errors on a logarithmic scale. The plot contrasts the error distributions between normal and anomalous data. The normal data, represented in blue, displays a tight interquartile range with fewer and lower outliers, indicating consistent low-error reconstructions. Anomalous data, shown in red, has a wider interquartile range and significant outliers, reflecting the expected higher reconstruction errors for deviations from the norm.

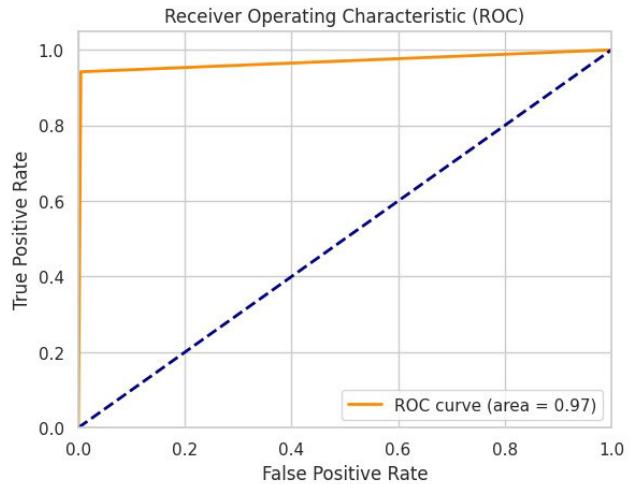


FIGURE 11. ROC curve illustrates the diagnostic ability of the LSTM-AE model as its discrimination threshold is varied. The TPR is plotted against the FPR, with the AUC quantifying the ability of the model to distinguish between normal and anomalous conditions. An AUC of 0.97 indicates a high level of separability, which implies that the model has a 97% chance of correct differentiation between anomaly and normal classes.

The whiskers, which indicate the range of the data excluding outliers, extend only slightly from the box, and have a few outliers, represented by individual points above the upper whisker.

By contrast, the anomaly data, shown in red, have a box with a higher median, demonstrating that anomalies tend to result in higher reconstruction errors. The IQR is broader, reflecting greater variability in the reconstruction errors for anomalies. This phenomenon is consistent with the expectation that anomalies will deviate from the normal pattern and thus be difficult to reconstruct accurately. The whiskers of the anomaly box plot stretch further from the IQR and have many outliers, indicating that some anomalies produce significantly higher errors.

When comparing Figure 10 with Figure 9, the box plot provides a different perspective. Although the histograms in Figure 9 depict the distribution of errors across the entire range, the box plot in Figure 10 emphasizes the central tendency and variability. The log scale highlights the differences between normal and anomalous data more prominently than a linear scale, particularly for data with a wide range of values.

The insights from Figure 10 affirm the findings from Figure 9, which provide evidence that the reconstruction error is a viable metric for distinguishing between normal and anomalies conditions. Additionally, the presence of outliers in the anomaly data emphasizes the necessity for robust threshold setting in the anomaly detection process.

Subsequently, we evaluated LSTM-AE model performance, particularly its ability to discriminate between normal and anomalous acoustic signals. Figures 11 and 12 show the representations of the discriminate ability of the LSTM-AE.

In this study, we utilized ROC curves to evaluate the performance of our model across various classification

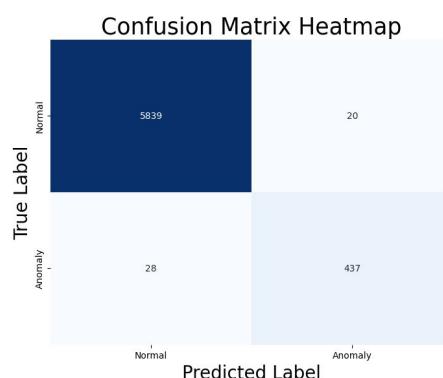


FIGURE 12. Confusion matrix provides a visual and quantitative representation of LSTM-AE classification performance. The matrix contrasts the number of true TP, TN, FP, and FN. In this case, the model correctly identified 5,831 normal samples and 438 anomalies, and only misclassified 28 normal samples as anomalies and failed to detect 20 anomalies.

thresholds. As shown in Figure 11, the ROC curve conveys a comprehensive perspective on the classification of the method prowess at various thresholds. With an AUC score of 97.0%, the model demonstrates an excellent probability of differentiating between the two classes, which is indicative of its robustness in anomaly detection. For the results presented in Table 3 and Figure 11, a specific threshold 3.0 was chosen to balance the trade-off between sensitivity and specificity effectively. This value was found to optimize the trade-off between TP and FP, leading to a balanced performance as shown in the ROC analysis. The corresponding performance metrics at this threshold are presented in Table 3, which include Accuracy, Recall, Precision and F1-score.

Complementing the ROC analysis, Figure 12 depicts the confusion matrix heatmap, which substantiates the

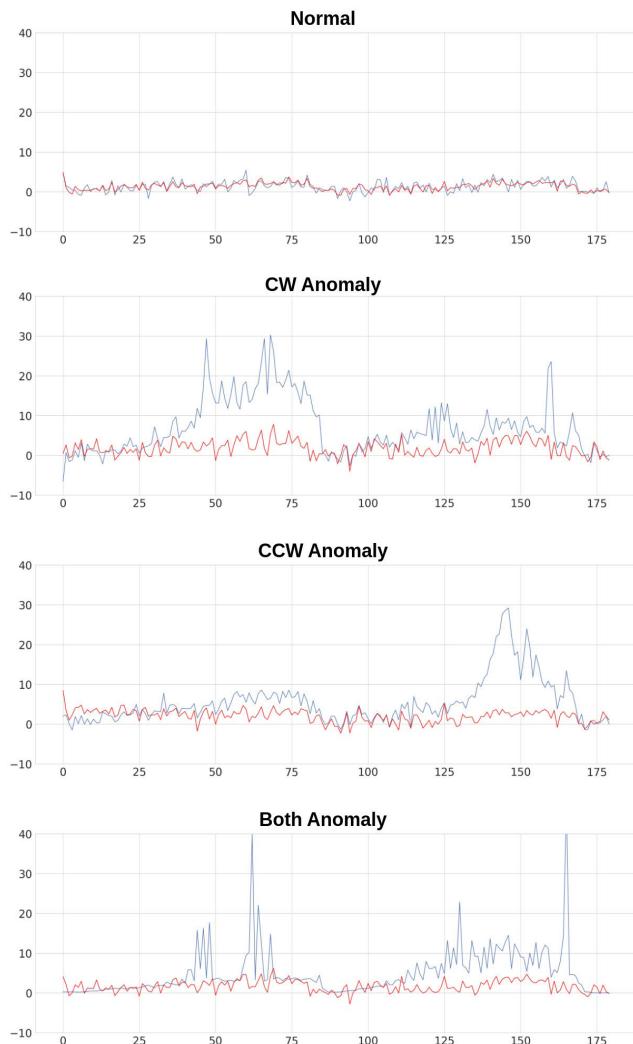


FIGURE 13. Input and output waveform comparison of LSTM-AE. This figure illustrates the reconstruction performance with the original input (blue) and the output of the LSTM-AE (red). The ‘normal’ plot indicates a close match between input and output, signifying accurate reconstruction. The ‘CW anomaly’ and ‘CCW anomaly’ plots reveal distinctive spikes in the output, which diverges from the input and are indicative of the anomalies the model has learned to reconstruct. The ‘both anomaly’ plot shows the response of LSTM-AE to a combination of anomalies, with marked deviations from the input, highlighting the sensitivity to multiple anomaly types.

precision of the method. The heatmap vividly contrasts true and false predictions, showing accurate classification while minimizing misclassification. The prevalence of true positive and true negative outcomes highlights the acumen of the method in accurately detecting anomalies.

These figures validate a convincing picture of the aptitude of the proposed method in the anomaly detection domain, highlighting its potential utility in real-world monitoring systems in which the cost of misdiagnosis can be substantial.

3) VISUALIZATION FOR ANOMALY DETECTION

Figure 13 through Figure 15 illustrate the ability of the LSTM-AE to discern and reconstruct normal and anomalous

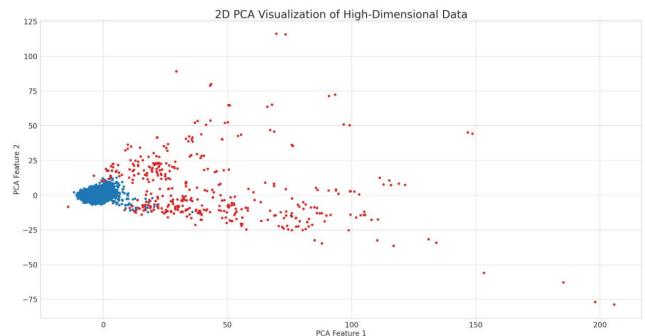


FIGURE 14. This figure provides a 2D representation of the high-dimensional latent space after PCA transformation. The blue points represent the normal data, closely clustered, reflecting their similarity in the latent space. By contrast, the red points, denoting anomalies, are dispersed, indicating the capability of the LSTM-AE to differentiate between normal and anomalous behavior based on the learned latent features.

acoustic patterns. Figure 13 presents a comparison of the input waveforms against the reconstructed outputs of LSTM-AE, with the red lines representing the output. A visual inspection reveals the proficiency in reconstructing normal patterns, as evidenced by the overlapping lines. However, the divergence in the anomaly cases characterized by spikes in the reconstructed output confirms the ability of the LSTM-AE to detect and signify deviations from normal waveform.

Figure 14 simplifies the high-dimensional latent space into a 2D plane using PCA, which demonstrates a clear separation between normal and anomalous data points. The dense clustering of normal data points indicates a strong similarity within normal conditions, whereas the scattered anomaly points reflect the recognition of varied anomalous states.

Advancing to a 3D visualization in Figure 15, we observe a stark contrast between data classes. The additional dimension reveals subtleties not visible in 2D, which provide a comprehensive understanding of latent space dynamics. The spatial separation between the normal and anomaly points affirms the effectiveness of LSTM-AE in encoding and distinguishes complex acoustic signatures.

In the analysis presented in Figure 15, the three principal components (PC1, PC2, and PC3) derived from PCA provide a multidimensional perspective on the latent features of the dataset. Here, we interpret the significance of each axis within the context of anomaly detection:

- PCA Feature 1 (PC1), represented on the x-axis, correlates with the extent of CW anomalies. As the value along this axis increases, it indicates an escalation in the severity of CW-related faults.
- PCA Feature 2 (PC2), denoted by the y-axis, exhibits an inverse relationship with CCW faults. Values descending into the negative spectrum indicate an intensification of CCW anomalies.
- PCA Feature 3 (PC3), depicted on the z-axis, presents a combined degree of CW and CCW faults, representing the overall severity of the anomalies.

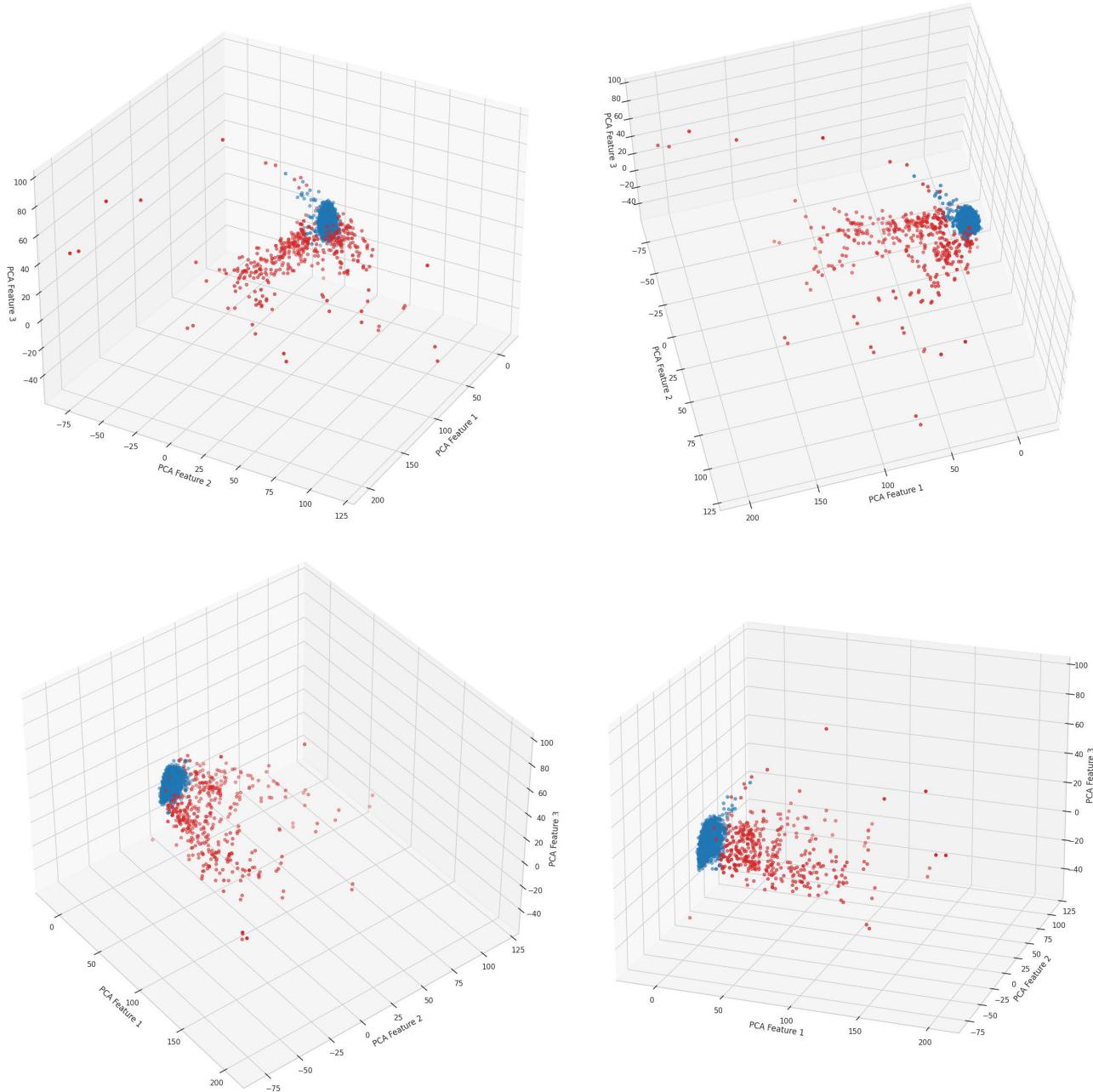


FIGURE 15. Expanding on Figure 14, this figure displays a 3D scatter plot of the latent space, where the PCA features afford a nuanced view of the data distribution. The depth added by the third dimension separates the normal (blue) and anomaly (red) data points, providing additional insight into the underlying data structure. This 3D perspective accentuates the distinction between the classes, showcasing the discriminative power in a high-dimensional feature space.

These axes collectively establish a diagnostic space in which the location of a data point within the 3D scatter plot can elucidate not only the presence of a fault but also its degree. A novel data sample position in this 3D distribution provides critical insights into the type of fault (CW or CCW), its magnitude, and the potential interplay between various fault types.

The 3D visualization of the latent space reveals clusters and patterns that are not discernible in lower-dimensional

representations. For instance, a dense congregation of points near the origin suggests normal conditions, whereas points that diverge considerably from this cluster indicate anomalies. The spatial distribution of these anomalies in relation to the defined axes (PC1, PC2, and PC3) provides a nuanced understanding of their characteristics.

By mapping new data points into this latent space, we can leverage the PCA features to make informed predictions regarding the motor condition. The clear demarcation of

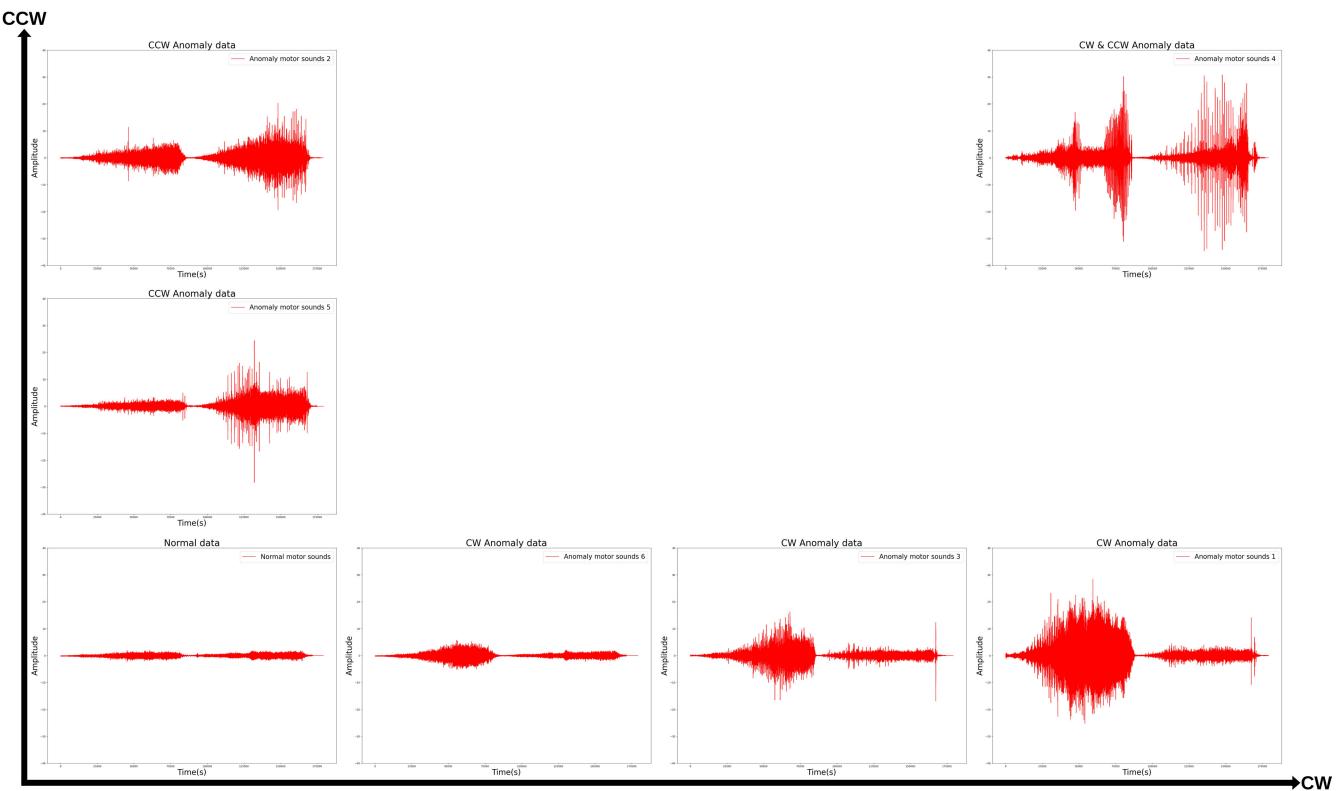


FIGURE 16. Classification of acoustic signals in latent space. This visualization categorizes the original raw sound signals by type within the latent space framework. The horizontal axis represents CW rotation anomalies, whereas the vertical axis corresponds to CCW rotation anomalies.

normal and anomalous regions within the space is a guide for identifying the state of the system, providing a valuable tool for pre-emptive maintenance and fault diagnosis.

The insights gleaned from the 3D latent space visualization enhance the utility of LSTM-AE as a diagnostic tool. This analysis method contributes to the robustness and interpretability of anomaly detection, providing a sophisticated approach to monitoring and maintaining system integrity in various applications.

In addition, visualization using PCA not only corroborates the theoretical capabilities of LSTM-AE, but also addresses several important limitations in anomaly detection tasks. Figure 13 shows the reconstructed graph of MFCC and DFMT of LSTM-AE, which requires careful manual adjustment of the threshold to distinguish between normal and defective data based on amplitude strength. This process is subjective and prone to error, particularly when intensity differences are subtle. PCA eliminates the need for these manual adjustments, providing a more objective and automated approach. PCA also normalizes the data, enabling it to detect anomalies based on underlying patterns rather than their degree of intensity. This normalization ensures a more robust anomaly detection process by reducing bias that can occur at different degrees of intensity. PCA focuses on principal components to capture essential variations and patterns in the data, enhancing the detection of anomalies that may

not be apparent in the time domain or envelope graph. Finally, certain data may exhibit defect shapes with low amplitude, which may be missed by simple thresholding of the envelope graph. These anomalies may fall below the threshold and be misclassified as normal. PCA solves this problem by analyzing key components of the data to detect anomalies based on unique features and patterns, regardless of their intensity. In summary, the PCA approach automates the anomaly detection process, reducing dependence on subjective threshold settings and improving the reliability of detection.

Figure 16 shows the classification of raw data signals based on their distribution, as depicted in Figure 15. The horizontal axis signifies the anomalies associated with CW rotations, whereas the vertical axis denotes anomalies related to CCW rotations. The origin point, where both axes intersect, represents the baseline of the original data.

As we move along the horizontal axis, an increase indicates a progressive deviation from the norm, showcasing examples of CW anomalies. In Figure 16, an upward movement along the y-axis exemplifies data samples that correspond to CCW anomalies. This phenomenon mirrors the representation in Figure 15 in which a negative decrease along the y-axis indicates a CCW anomaly. Data that exhibit anomalies in both CW and CCW rotations are displayed toward the upper regions of both axes, emphasizing their compounded

characteristics. Therefore, in the context of an acoustic data of EPS motor, we could differentiate between normal and anomaly conditions. Within the anomaly category, we can classify the data based on the type of rotational anomaly CW or CCW.

The insights from Figure 16 provide a comprehensive understanding of anomalies within the acoustic data of EPS motors. This visualization can not only be used as a diagnostic tool to identify the type of rotational fault but also highlights the ability of the method to isolate and identify specific types of anomalies.

IV. CONCLUSION

In this study, we introduced a methodology for detecting acoustic anomalies in EPS electronic motors, featuring an advanced feature extraction technique that efficiently reduces feature dimensionality while maintaining the original raw sound waveform. The proposed DFMT method can preserve essential waveform features, which are pivotal for precise anomaly detection. To ensure that vital low-frequency information is not overlooked, we combined DFMT with MFCC to enhance the feature extraction process. This combined strategy incorporates the distinct advantages of both DFMT and capacity of MFCC-DFMT for capturing detailed waveform and effectiveness of MFCC in summarizing frequency domain information. We used LSTM-AE for classifying data as normal or anomalous, based on reconstruction error metrics, highlighting the efficacy of the proposed feature extraction approach. Additionally, we performed PCA on the latent space values from the LSTM-AE, enabling 3D visualizations that swiftly elucidated the distribution of anomalies and degree. These visual representations provide considerable insights into motor conditions, illustrating the presence and degree of anomalies in the acoustic signals of the EPS electronic motor.

This study provides enhanced accuracy and insightful analysis of motor conditions for acoustic anomaly detection. The proposed methodology can accurately identify and classify anomalies. The experimental results showed that the proposed method achieved an accuracy of 99.2%, a recall of 94.0%, a precision of 95.6%, and an F1-score of 94.7%. This results highlight potential for applications in predictive maintenance and quality control in manufacturing processes. By integrating DFMT with MFCC and leveraging the classification capabilities of LSTM-AE alongside PCA for visualization, this study not only progresses the theoretical understanding of anomaly detection but also lays the groundwork for subsequent exploration.

In Future studies, the performance of the method should be studied in various settings with different noise levels and operational conditions to evaluate its adaptability and robustness. Moreover, real-time data processing and automated threshold-determination mechanism should be incorporated in the proposed method to expand its applicability to dynamic operational systems, enhancing its utility in real-world scenarios.

REFERENCES

- [1] L. W. Alabe, K. Kea, Y. Han, Y. J. Min, and T. Kim, "A deep learning approach to detect anomalies in an electric power steering system," *Sensors*, vol. 22, no. 22, p. 8981, Nov. 2022.
- [2] J.-H. Kim and J.-B. Song, "Control logic for an electric power steering system using assist motor," *Mechatronics*, vol. 12, no. 3, pp. 447–459, Apr. 2002.
- [3] A. A. Badawy, F. Bolourchi, and S. K. Gaut, "E steer system redefines steering technology," *Automot. Syst.*, pp. 15–18, 1997.
- [4] W.-C. Lin and Y. A. Ghoneim, "Model-based fault diagnosis and prognosis for electric power steering systems," in *Proc. IEEE Int. Conf. Prognostics Health Manage. (ICPHM)*, Jun. 2016, pp. 1–8.
- [5] A. T. Zaremba, M. K. Liubakka, and R. M. Stuntz, "Control and steering feel issues in the design of an electric power steering system," in *Proc. Amer. Control Conf. (ACC)*, vol. 2, Jun. 1998, pp. 36–40.
- [6] S. Gao, Y. Ren, Y. Zhang, and T. Li, "Fault diagnosis of rolling bearings based on improved energy entropy and fault location of triangulation of amplitude attenuation outer raceway," *Measurement*, vol. 185, Nov. 2021, Art. no. 109974.
- [7] F. Zhang, J. Huang, F. Chu, and L. Cui, "Mechanism and method for the full-scale quantitative diagnosis of ball bearings with an inner race fault," *J. Sound Vibrat.*, vol. 488, Dec. 2020, Art. no. 115641.
- [8] F. Zhang, J. Huang, F. Chu, and L. Cui, "Mechanism and method for outer raceway defect localization of ball bearings," *IEEE Access*, vol. 8, pp. 4351–4360, 2020.
- [9] X. Chen, R. Yang, Y. Xue, M. Huang, R. Ferrero, and Z. Wang, "Deep transfer learning for bearing fault diagnosis: A systematic review since 2016," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–21, 2023.
- [10] K. Suefusa, T. Nishida, H. Purohit, R. Tanabe, T. Endo, and Y. Kawaguchi, "Anomalous sound detection based on interpolation deep neural network," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 271–275.
- [11] Y. Koizumi, S. Saito, H. Uematsu, Y. Kawachi, and N. Harada, "Unsupervised detection of anomalous sound based on deep learning and the Neyman–Pearson lemma," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 27, no. 1, pp. 212–224, Jan. 2019.
- [12] Y. Koizumi, S. Saito, H. Uematsu, N. Harada, and K. Imoto, "ToyADMOS: A dataset of miniature-machine operating sounds for anomalous sound detection," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, Oct. 2019, pp. 313–317.
- [13] D. Y. Oh and I. D. Yun, "Residual error based anomaly detection using auto-encoder in SMD machine sound," *Sensors*, vol. 18, no. 5, p. 1308, Apr. 2018.
- [14] H. Yun, H. Kim, Y. H. Jeong, and M. B. G. Jun, "Autoencoder-based anomaly detection of industrial robot arm using stethoscope based internal sound sensor," *J. Intell. Manuf.*, vol. 34, no. 3, pp. 1427–1444, Mar. 2023.
- [15] Y. Koizumi, Y. Kawaguchi, K. Imoto, T. Nakamura, Y. Nikaido, R. Tanabe, H. Purohit, K. Suefusa, T. Endo, M. Yasuda, and N. Harada, "Description and discussion on DCASE2020 challenge task2: Unsupervised anomalous sound detection for machine condition monitoring," 2020, *arXiv:2006.05822*.
- [16] H. Zhang, Q. Zhu, J. Guan, H. Liu, F. Xiao, J. Tian, X. Mei, X. Liu, and W. Wang, "First-shot unsupervised anomalous sound detection with unknown anomalies estimated by metadata-assisted audio generation," 2023, *arXiv:2310.14173*.
- [17] S. Yan, H. Shao, Y. Xiao, B. Liu, and J. Wan, "Hybrid robust convolutional autoencoder for unsupervised anomaly detection of machine tools under noises," *Robot. Computer-Integrated Manuf.*, vol. 79, Feb. 2023, Art. no. 102441.
- [18] J. Guan, Y. Liu, Q. Kong, F. Xiao, Q. Zhu, J. Tian, and W. Wang, "Transformer-based autoencoder with ID constraint for unsupervised anomalous sound detection," *EURASIP J. Audio, Speech, Music Process.*, vol. 2023, no. 1, Oct. 2023, Art. no. 42.
- [19] N. Harada, D. Niizumi, Y. Ohishi, D. Takeuchi, and M. Yasuda, "First-shot anomaly sound detection for machine condition monitoring: A domain generalization baseline," 2023, *arXiv:2303.00455*.
- [20] K. Dohi, K. Imoto, N. Harada, D. Niizumi, Y. Koizumi, T. Nishida, H. Purohit, R. Tanabe, T. Endo, and Y. Kawaguchi, "Description and discussion on DCASE 2023 challenge task 2: First-shot unsupervised anomalous sound detection for machine condition monitoring," 2023, *arXiv:2305.07828*.

- [21] 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.
- [22] Z. Huang, B. Zhang, G. Hu, L. Li, Y. Xu, and Y. Jin, "Enhancing unsupervised anomaly detection with score-guided network," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jun. 7, 2023, pp. 1–16, doi: 10.1109/TNNLS.2023.3281501.
- [23] L. Ruff, J. R. Kauffmann, R. A. Vandermeulen, G. Montavon, W. Samek, M. Kloft, T. G. Dietterich, and K.-R. Müller, "A unifying review of deep and shallow anomaly detection," *Proc. IEEE*, vol. 109, no. 5, pp. 756–795, May 2021.
- [24] D. A. Reynolds, "Gaussian mixture models," *Encyclopedia Biometrics*, vol. 741, 2009, Art. no. 659663.
- [25] B. Zong, Q. Song, M. R. Min, W. Cheng, C. Lumezanu, D. Cho, and H. Chen, "Deep autoencoding Gaussian mixture model for unsupervised anomaly detection," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2018.
- [26] W. Hu, J. Gao, B. Li, O. Wu, J. Du, and S. Maybank, "Anomaly detection using local kernel density estimation and context-based regression," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 2, pp. 218–233, Feb. 2020.
- [27] L. Bergman, N. Cohen, and Y. Hoshen, "Deep nearest neighbor anomaly detection," 2020, *arXiv:2002.10445*.
- [28] H. S. Emadi and S. M. Mazinani, "A novel anomaly detection algorithm using DBSCAN and SVM in wireless sensor networks," *Wireless Pers. Commun.*, vol. 98, pp. 2025–2035, Jan. 2018.
- [29] J. An and S. Cho, "Variational autoencoder based anomaly detection using reconstruction probability," *Special Lect. IE*, vol. 2, no. 1, pp. 1–18, 2015.
- [30] Z. Chen, C. K. Yeo, B. S. Lee, and C. T. Lau, "Autoencoder-based network anomaly detection," in *Proc. Wireless Telecommun. Symp. (WTS)*, Apr. 2018, pp. 1–5.
- [31] J. K. Chow, Z. Su, J. Wu, P. S. Tan, X. Mao, and Y. H. Wang, "Anomaly detection of defects on concrete structures with the convolutional autoencoder," *Adv. Eng. Informat.*, vol. 45, Aug. 2020, Art. no. 101105.
- [32] S. Perez-Castanos, J. Naranjo-Alcazar, P. Zuccarello, and M. Cobos, "Anomalous sound detection using unsupervised and semi-supervised autoencoders and gammatone audio representation," 2020, *arXiv:2006.15321*.
- [33] S. Abbasi, M. Famouri, M. J. Shafiee, and A. Wong, "OutlierNets: Highly compact deep autoencoder network architectures for on-device acoustic anomaly detection," *Sensors*, vol. 21, no. 14, p. 4805, 2021.
- [34] M. S. Elsayed, N.-A. Le-Khac, S. Dev, and A. D. Jurcut, "Network anomaly detection using LSTM based autoencoder," in *Proc. 16th ACM Symp. QoS Secur. Wireless Mobile Netw.*, 2020, pp. 37–45.
- [35] M. Sakurada and T. Yairi, "Anomaly detection using autoencoders with nonlinear dimensionality reduction," in *Proc. 2nd Workshop Mach. Learn. Sensory Data Anal. (MLSDA)*, Dec. 2014, pp. 4–11.
- [36] Z. K. Abdul and A. K. Al-Talabani, "Mel frequency cepstral coefficient and its applications: A review," *IEEE Access*, vol. 10, pp. 122136–122158, 2022.
- [37] J. Lin, "Feature extraction of machine sound using wavelet and its application in fault diagnosis," *NDT E Int.*, vol. 34, no. 1, pp. 25–30, Jan. 2001.
- [38] M. K. Singh, S. Kumar, and D. Nandan, "Faulty voice diagnosis of automotive gearbox based on acoustic feature extraction and classification technique," *J. Eng. Res.*, vol. 11, no. 2, Jun. 2023, Art. no. 100051.
- [39] F. V. Newamondo and T. Marwala, "Faults detection using Gaussian mixture models, mel-frequency cepstral coefficients and kurtosis," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Oct. 2006, pp. 290–295.
- [40] Q. Jiang, F. Chang, and B. Sheng, "Bearing fault classification based on convolutional neural network in noise environment," *IEEE Access*, vol. 7, pp. 69795–69807, 2019.
- [41] Z. K. Abdul, A. K. Al-Talabani, and D. O. Ramadan, "A hybrid temporal feature for gear fault diagnosis using the long short term memory," *IEEE Sensors J.*, vol. 20, no. 23, pp. 14444–14452, Dec. 2020.
- [42] P. Yao, J. Wang, F. Zhang, W. Li, S. Lv, M. Jiang, and L. Jia, "Intelligent rolling bearing imbalanced fault diagnosis based on mel-frequency cepstrum coefficient and convolutional neural networks," *Measurement*, vol. 205, Dec. 2022, Art. no. 112143.
- [43] Y. Zhao, N. Zhang, Z. Zhang, and X. Xu, "Bearing fault diagnosis based on mel frequency cepstrum coefficient and deformable space-frequency attention network," *IEEE Access*, vol. 11, pp. 34407–34420, 2023.
- [44] V. Verma, A. Benjwal, A. Chhabra, S. K. Singh, S. Kumar, B. B. Gupta, V. Arya, and K. T. Chui, "A novel hybrid model integrating MFCC and acoustic parameters for voice disorder detection," *Sci. Rep.*, vol. 13, no. 1, p. 22719, Dec. 2023.
- [45] S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. A. Liu, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. Woodland, *The HTK Book*, vol. 3. Cambridge, U.K.: Cambridge Univ. Press, Engineering Department, 2002, p. 175.
- [46] H. Torabi, S. L. Mirtaheri, and S. Greco, "Practical autoencoder based anomaly detection by using vector reconstruction error," *Cybersecurity*, vol. 6, no. 1, p. 1, Jan. 2023.
- [47] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educ. Psychol.*, vol. 24, no. 6, pp. 417–441, Sep. 1933.
- [48] Y. Tagawa, R. Maskeliūnas, and R. Damasevičius, "Acoustic anomaly detection of mechanical failures in noisy real-life factory environments," *Electronics*, vol. 10, no. 19, p. 2329, Sep. 2021.
- [49] X. Xu, H. Wen, H. Lin, Z. Li, and C. Huang, "Online detection method for variable load conditions and anomalous sound of hydro turbines using correlation analysis and PCA-adaptive-K-means," *Measurement*, vol. 224, Jan. 2024, Art. no. 113846.



EUN SUN YUN received the bachelor's degree in mathematics from Chungnam National University, Daejeon, South Korea, in 2017, and the bachelor's degree in computer science from Aix-Marseille University, Marseille, France, in 2021, and the master's degree in data and HPC from the University of Science and Technology (UST), in 2024, where she is currently pursuing the Ph.D. degree in data and HPC. She is also a Student Researcher with Korea Institute of Science and Technology Information (KISTI). Her research interests include data clustering/pattern recognition using machine learning/deep learning algorithms.



MINJOONG JEONG received the Ph.D. degree in frontier science (mechanical design) from Tokyo University, Tokyo, Japan, in 2004. He is currently a Principal Researcher with the National Supercomputing Center, Korea Institute of Science and Technology Information (KISTI). He is also a Professor with the University of Science and Technology (UST), South Korea. His research interests include multiobjective/multicriterion optimization, data clustering/pattern recognition using machine learning, and evolutionary algorithms for optimization.