SINERGI Vol. 30, No. 1, February 2026: 173-184
http://publikasi.mercubuana.ac.id/index.php/sinergi
http://doi.org/10.22441/sinergi.2026.1.016

Towards enhanced acoustic fan booster damage detection: a comparative study of feature-based and machine learning approaches

Rikko Putra Youlia[1], Dedik Romahadi[1,2,*], Aberham Genetu Feleke[2], Irfan Evi Nugroho[1], Alina Alina[3]
[1]Department of Mechanical Engineering, Faculty of Engineering, Universitas Mercu Buana, Indonesia
[2]School of Mechanical Engineering, Beijing Institute of Technology, China
[3]Department of Information Systems, Universitas Terbuka, Indonesia

Abstract
Machine failure detection frequently uses non-destructive monitoring techniques such as vibration analysis. Although vibration analysis can identify machine degradation, the apparatus is often costly and necessitates specialist knowledge. Additionally, many existing methods in audio classification rely on characteristics represented as pictures or vectors, which increases computational complexity. In contrast, this research introduces a novel method that substitutes vibration data with a singular numerical feature derived from audio signals, addressing both cost and complexity issues. Our objective is to develop a rapid and precise audio-based method for detecting machine damage. The acoustic signals from the machine apparatus were classified into three categories: normal, belt damage, and combined belt and bearing defect. The data processing technique involved lowering the sample rate and segmenting the data to improve computational efficiency and classification performance. We use the Welch method and appropriate statistical techniques to analyze Power Spectral Density (PSD). The performance of seven classifier models, KNN, LDA, SVM, NB, ANN, RF, and DT, was evaluated using accuracy, precision, sensitivity, specificity, and F-score. LDA achieved the highest accuracy at 92.83%, followed by ANN (92.75%), NB (92.74%), and DT (92.34%). These models outperformed KNN (89.90%) and RF (89.40%), with SVM recording the lowest accuracy at 85.40%. LDA was highly effective, achieving the highest accuracy with a single average PSD-type feature, showcasing its robustness in machine defect diagnosis. Compared to previous methods, this approach simplifies feature extraction, reduces computational demands, and maintains high diagnostic performance, providing notable benefits in terms of effectiveness and precision.

Keywords:
Acoustic detection;
Fault detection;
Machine learning;
Power spectral density;
Welch method;

Article History:
Received: May 6, 2025
Revised: August 27, 2025
Accepted: October 2, 2025
Published: January 16, 2026

Corresponding Author:
Dedik Romahadi
Department of Mechanical Engineering, Faculty of Engineering, Universitas Mercu Buana, Indonesia
School of Mechanical Engineering, Beijing Institute of Technology, China
Email:

INTRODUCTION

Rotating machinery constitutes the foundation of the manufacturing industry. Any damage or malfunction in these components during operation might severely affect the system's performance [1]. Consequently, it is crucial to develop effective and accurate methods for diagnosing and categorizing problems to improve machine reliability and substantially decrease operational and maintenance costs [2], [3]. The Electric Power Research Institute reports that bearing components constitute an astonishing 41% of failures in rotating machinery. Consequently, machines with rotating components have emerged as a significant focus of research, with an opportunity to employ artificial

intelligence methodologies for high-performance maintenance practices [4][5].

Cost-effective maintenance of industrial machinery often involves the use of spinning parts to perform particular tasks. This system consists of gears, rotor shafts, rotors linked by fasteners, and bearings. Rolling bearings play a crucial role in industrial machinery by enabling the rotation of components with minimal friction. [6]. As a result, previous studies have thoroughly explored rolling bearings. The primary causes of problems are generally localized wear on the inner race (BPFI), the outer race (BPFO), or the balls (BSF) of the bearing. Defects on the rolling surface may include cracks, voids, and fragments. The most common type of damage occurs when a ball contacts a slight imperfection. A crash between a rolling component and a flaw generates a bump that causes high-frequency vibrations throughout the structure [7]. According to Lou et al. [8], when capacitive contact occurs, applying a voltage exceeding 73% of the total breakdown voltage causes metal erosion within the bearing and damages its lubricating layer. As a result, the capacitive contact transitions to a resistive contact, which helps reduce electrical erosion while accelerating oxidation.

The essential components of a condition monitoring system are defect identification and severity assessment [9]. This is vital, especially in critical systems necessitating continuous operation. Various approaches have been proposed to track and assess rolling element bearings, given their vital role in rotating machinery and the severe consequences of undetected damage [10]. Chen et al. [11] emphasized that the sampling frequency significantly influences fault diagnosis performance, underscoring the importance of data-acquisition settings. Most diagnostic methods aim to identify faults and assess their severity, essential for uninterrupted functioning in critical systems [12][13]. The ball-passing frequency, a typical defect frequency, arises from periodic impacts determined by the bearing's shape and rotational rate. Despite noise from other components, which can obscure or distort signals, researchers have consistently highlighted the FFT's reliability and well-established capabilities for diagnosing bearing faults [14]. While FFT-based analysis is robust, its implementation can be costly due to the specialized equipment required. To address this limitation, a promising alternative is to substitute vibration signals with audio signals captured by a mobile phone. Vibration signals are typically measured using specialized sensors such as accelerometers or piezoelectric transducers, which detect minute changes in the vibrations of machine components. These signals are then processed using signal conditioning equipment and data-acquisition systems to extract meaningful fault-diagnosis information.

Non-destructive monitoring techniques, specifically vibration and acoustics, are the predominant methods for diagnosing machine failures [15]. Recent review studies have been undertaken to elucidate common defects in rolling bearings, and analytical methodologies, including the research presented by [16]. Other reviews focus on identifying the most reliable indicators to improve defect-detection accuracy. Kuncan [17] evaluated different machine learning techniques, including Support Vector Machine (SVM), k-nearest neighbors (KNN), Artificial Neural Network (ANN), Linear Discriminant Analysis (LDA), Logistic Regression (LR), and Gray Relational Analysis (GRA), for detecting bearing faults. The GRA approach achieved a 100% success rate across the complete set of four signal datasets used for bearing signal classification, while SVM and LDA also yielded commendable classification results. Hosseini et al. [18] proposed an algorithm that emphasizes the benefits of employing a Projection Recurrent Neural Network within an SVM framework. The findings suggest that the suggested model exceeds the recently developed models in fault detection accuracy. An extensive analysis of the most commonly utilized ML models for bearing defect detection is provided to ensure appropriate deployment [19].

Synthesizing this literature reveals that while vibration and acoustics-based monitoring combined with advanced machine learning (ML) algorithms have shown promise in machine fault detection, several key gaps persist. Many studies either focus exclusively on a single signal type or fail to thoroughly assess the effectiveness of specific feature-extraction techniques, thereby limiting their practical applicability. Additionally, while significant research has been conducted on the use of Power Spectral Density (PSD) features and ML models for fault classification, these efforts often rely on multiple features or complex models like deep learning, which are computationally intensive and not always optimal for real-time applications. A notable gap in the literature is the lack of studies on using a single, computationally efficient PSD feature—particularly one derived from the Welch method’s statistical value—to identify and classify machine damage types. This is significant because such a streamlined approach simplifies feature extraction and enables more efficient industrial deployment without sacrificing diagnostic performance.

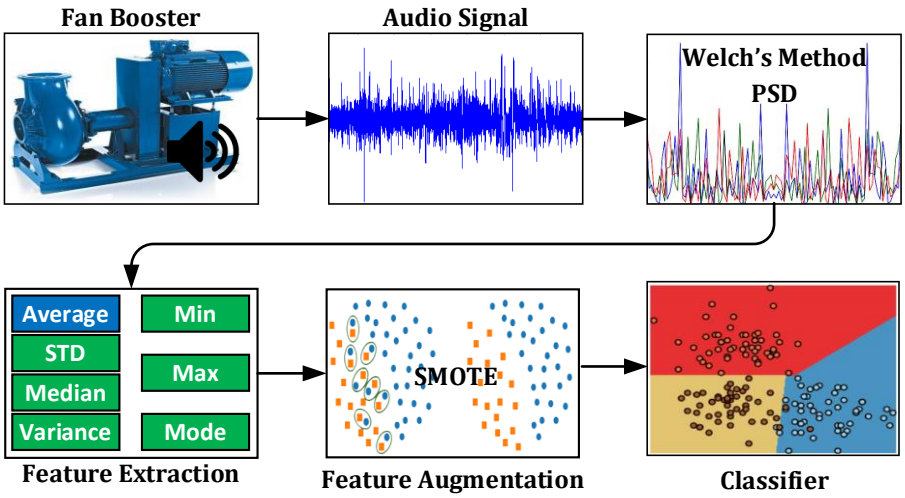

Figure 1. The overview of the machine acoustic signal classification process

Furthermore, despite the increasing popularity of deep learning models in the field, simpler models such as LDA or SVM have shown competitive or superior performance in many cases, particularly when the dataset is limited or the problem structure does not warrant the complexity of deep networks. This suggests that simpler, more efficient models can be as effective in specific diagnostic tasks. Our research addresses this gap by applying the Welch method's statistical value as a single feature extraction technique, evaluated using a 5-fold cross-validation scheme to ensure robust system performance. This approach leverages the strengths of audio-based monitoring and the proven power of simpler ML models, offering both scientific and practical contributions to machine fault diagnosis, with significant improvements in computational efficiency and cost-effectiveness over traditional vibration-based systems.

MATERIAL AND METHOD

Our research aimed to develop a robust system for accurately categorizing various types of machine damage using acoustic pattern analysis. The procedure for identifying damage typically includes five key steps: data collection, sound preprocessing, feature extraction, feature selection, and classification. The approach for defect identification is shown in Figure 1. The machine's acoustic data undergoes a preprocessing step to reduce computational time and ensure suitability for further analysis. During the second stage, the data within each segment is transformed into PSD using the Welch algorithm, with adjustments made to account for statistical variations. Next, the Synthetic Minority Over-sampling Technique (SMOTE) is used to create supplementary samples. The feature selection process then minimizes the influence of unnecessary factors. Since only a limited number of features are available, each feature's contribution to the classification performance is thoroughly evaluated.

Audio Acquisition & Preprocessing

The data processing tasks were executed on a personal computer powered by Windows 11 64-bit, equipped with an RTX 4080 Super graphics card and a 14th-generation Intel Core i9 processor. In this experiment, the graphics card was not utilized. The analysis was carried out in MATLAB (version 2023b) [20], utilizing the Welch method and classifiers from the MATLAB toolbox to retrieve machine sound data from the dataset. Table 1 summarizes the number of samples used for classification. These numbers correspond to the number of distinct chunks or segments extracted from longer recordings. Each chunk represents a specific portion of a continuous recording, enabling more manageable, detailed analysis during classification. The audio recordings were captured using the built-in recording function of a Vivo V21 mobile device at a sampling rate of 48 kHz, with a maximum bitrate of 127 kbps and support for stereo channels. During data collection, the environment was monitored for background noise, which was kept to a minimum to ensure clear, accurate recordings. However, occasional ambient sounds, such as background chatter or environmental noise, were present at low levels but were considered negligible in the analysis. The recordings were saved in the m4a format. The fan booster apparatus used for these recordings is shown in Figure 2, with its specifications provided in Tables 2 and 3.

Throughout the data collection stage, we gathered sound samples from three different machine scenarios involving fan boosters, which were categorized as follows: (a) the vehicle belt is in standard condition with properly adjusted tension, (b) the van belt is loosely secured with incorrect tension adjustment, and (c) the serpentine belt is inadequately fastened, and the bearings exhibit defects, that is a large bearing clearance, which can lead to excessive play between the bearing components. This increased clearance can result in vibrations, noise, and reduced load-carrying capacity, ultimately affecting the performance and lifespan of the machinery. Next, the audio data was downsampled from 48 kHz to 22.05 kHz to enhance computational efficiency. We also converted the audio channel configuration from stereo to mono. Initially, each category had over 3 hours of audio recordings. To manage this, we segmented the audio into non-overlapping 5-second sections. The audio data had a sampling rate of 22.05 kHz, a bit rate of 22.05 kbps, was recorded in a mono-channel configuration, and was saved in the m4a format.

Audio Signal Feature Extraction & Selection

The Welch algorithm computes a modified frequency-domain representation for each data segment and then averages these representations to estimate the PSD [21]. Constructed from PSD estimates of different time-series segments, the modified periodogram, produced using the Welch method, provides an uncorrelated estimate of the actual PSD. This averaging approach minimizes fluctuations. This study aims to estimate the PSD of machine audio signals using the Welch method. The Welch method applies the Hann window function to each data segment, which helps reduce spectral leakage. By calculating an adjusted periodogram for each segment and averaging the results, the approach provides a more accurate estimate of the PSD.

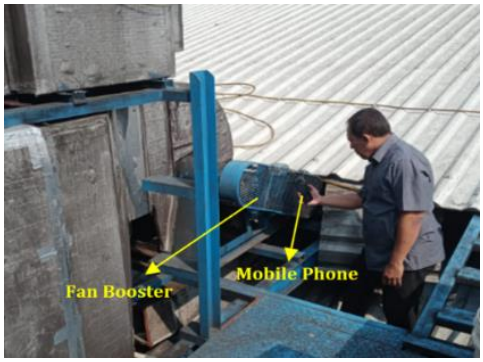

Figure 2. Data acquisition on the fan booster machine

Table 1. Number of samples

Classes	Number	Percentage
Belt	2350	34.18
Belt-Bearing	2275	33.09
Normal	2250	32.73

Table 2. Specifications of the fan assembly

No.	Description	Specification
1	Brand	Topindo Fan
2	Type	RSH 500
3	Serial Number	-
4	Fan Speed	1400 rpm
5	Power Supply	3 PH; 380/660 V; 50 Hz
6	Capacity	7800 m^3/h
7	Static Pressure	875 Pa (N/M^2)
8	Fan belt	Type V Belt B72

Table 3. Motor configuration

No.	Description	Specification
1	Type	1A1325-4
2	Serial Number	-
3	Power	50 Hz; 5.5 Kw
4	Power Input	380/660 V 11.84 / 6.82 A ~ 440 VA 11.84 A
5	Speed	1440 rpm
6	Efficiency	85%

The overlap between adjacent segments increases frequency resolution, while windowing mitigates data loss at segment boundaries, especially at edges.

The auditory data x and y, each comprising N samples, are partitioned into blocks $x^{(k)}$ and $y^{(k)}$ of length L. The relevant blocks in the estimated spectrum $\hat{S}_{xy}^{(k)}$ are presented in (1).

$$\hat{S}_{xy}^{(k)}(f) = \frac{2 \cdot \mathrm{F}\left(x^{(k)} \cdot w\right) \cdot conj\left(\mathrm{F}\left(y^{(k)} \cdot w\right)\right)}{fs \cdot \sum w_i^2} \quad (1)$$

The Fourier Transform F is applied to both signals, with a window function w used to mitigate edge effects and improve frequency resolution. The term $x^{(k)} \cdot w$ $y^{(k)} \cdot w$ indicates that the window function is applied to the k-th data samples of the signals before computing their respective Fourier Transforms. The conjugate of the Fourier Transform of $y^{(k)}$, denoted as $conj\left(\mathrm{F}\left(y^{(k)} \cdot w\right)\right)$, is then multiplied by the Fourier Transform of $x^{(k)}$, capturing the phase and amplitude relationships between the two signals in the frequency domain. The resulting product is divided by a normalization factor, the sum of the squared weights w_i^2, , which accounts for the influence of the window function, and by a scaling factor s, potentially related to the signal length. The cross-spectrum is computed by averaging the components in (2).

$$\hat{S}_{xy}(f)=\sum_{k=1}^{N/L}\frac{\hat{S}_{xy}^{(k)}}{N/L} \quad (2)$$

A key reason for averaging the PSD over multiple segments, as in the Welch method, is that it reduces the variance of the periodogram. The periodogram of a single segment often exhibits high variance due to noise and short-duration signals, leading to unreliable PSD estimates. By averaging across overlapping segments, the Welch method smooths out random fluctuations, yielding a more consistent and statistically significant estimate of the PSD. This approach reduces bias arising from spectral leakage and provides a more accurate representation of the underlying signal's power distribution across frequencies.

Following the Welch method, distinct 5-second intervals of EEG data were analyzed without overlap using a short-time Fourier transform to compute the PSD properties. Seven statistical measures, encompassing the mean, median, mode, standard deviation, variance, minimum, and maximum, were applied to generate PSD attributes for each interval. The most relevant feature was then identified by evaluating the computational significance of these measures and prioritizing outcomes based on the analytical results derived from their statistical properties.

$$x_i^{'}=x_i+r\times\left(z_{ij}-x_i\right) \quad (3)$$

Equation 3 illustrates how the SMOTE algorithm generates synthetic samples to address class imbalance in sample sizes [22]. In this method, a new synthetic sample $x_i^{'}$ is created by interpolating between an existing minority-class sample x_i and one of its randomly selected neighbors z_{ij}. The interpolation is controlled by a random factor r, which ranges from 0 to 1, specifying the location of the updated sample along the line segment linking x_i z_{ij} and.

Classification

We obtained multiple enhanced features after applying the SMOTE algorithm for data balancing. We conducted experiments with seven different models. These models were selected for their simplicity, fast computation, and widespread use in previous research.

1) k-nearest neighbors (KNN): In KNN, the classification is based on the majority vote of the k nearest neighbors to a data point [23]. The decision rule is mathematically expressed as:

$$\hat{y}=\text{mode}(y_1,y_2,\cdots,y_k) \quad (4)$$

Where y_i represents the class labels of the k nearest neighbors of a point.

2) Linear Discriminant Analysis (LDA): LDA is a method used for dimensionality reduction and classification [24]. It seeks to find a linear combination of features that best separates two or more classes. For each class i, the mean vector is defined as:

$$\mu_i=\frac{1}{N_i}\sum_{x_j\in C_i}x_j \quad (5)$$

Where N_i is the number of data points in class C_i, , and x_j are the data points in class C_i. The within-class scatter matrix measures the variance within each class:

$$S_W=\sum_{i=1}^{k}\sum_{x_j\in C_i}\left(x_j-\mu_i\right)\left(x_j-\mu_i\right)^T \quad (6)$$

Where k is the number of classes, and $\mu_i\,\varepsilon$ is the mean of the class C_i. The between-class scatter matrix measures the variance between the class means and the overall mean:

$$S_B=\sum_{i=1}^{k}N_i\left(\mu_i-\mu\right)\left(\mu_i-\mu\right)^T \quad (7)$$

Where μ is the overall mean of all data points, and N_i represents the quantity of data points in class C_i. To find the optimal projection:

$$S_W^{-1}S_B\upsilon=\lambda\upsilon \quad (8)$$

Where υ is the eigenvector and λ is the corresponding eigenvalue. The discriminant function that LDA uses to classify a new sample x is:

$$y=\arg\max_i\left(\mu_i^T S_W^{-1}x-\frac{1}{2}\mu_i^T S_W^{-1}\mu_i+\log P(C_i)\right) \quad (9)$$

Where $P(C_i)$ is the initial probability of the class C_i, and the other terms relate to the class-specific means and scatter matrices. The decision rule assigns the sample to the class with the highest value.

3) Support Vector Machine (SVM): SVM constructs a hyperplane to separate classes with the maximum margin [25]. The decision function is expressed as:

$$f(x) = \sum_{i=1}^{N} \alpha_i y_i K(x, x_i) + b \quad (10)$$

Where α_i are the Lagrange multipliers, y_i are the class labels, $K(x, x_i)$ is the kernel function, and b is the bias term.

4) Naive Bayes (NB): In Naive Bayes, the classification relies on Bayes' theorem, assuming that the features are conditionally independent given the class [26]. The rule is:

$$P(y|\ x) = \frac{P(x|\ y)P(y)}{P(x)} \quad (11)$$

Where $P(y|\ x)$ is the posterior probability, $P(x|\ y)$ is the likelihood, and $P(y)$ is the prior probability.

5) Artificial Neural Network (ANN): The output of an ANN is calculated using the weights and biases of the network and an activation function [27]. It is mathematically represented as:

$$y = f(Wx + b) \quad (12)$$

Where W represents the weights, x is the input vector, b is the bias term, and f is the activation function. The ANN is a medium-sized neural network with 25 units per hidden layer and ReLU activation. The parameter Lambda is set to 0, indicating that no regularization is applied to prevent overfitting. The Iteration limit is set to 1000, specifying the maximum number of iterations the network will perform during training. Additionally, the standardize option is enabled, meaning the input data is standardized to have a mean of zero and a standard deviation of one. These models were used to assess how well various attributes identified damage to the fan booster.

6) Random Forest (RF): Random Forest builds an ensemble of decision trees and uses majority voting to make predictions [28]. The decision rule is given by:

$$y = mode(T_1(x), T_2(x), \ldots, T_M(x)) \quad (13)$$

Where $T_M(x)$ represents the prediction of the m-th decision tree, and the mode is the majority vote from all trees.

7) Decision Tree (DT): In a decision tree, decisions are made by splitting the dataset based on feature thresholds [29]. The classification rule can be represented as:

$$y = f(x) \quad (14)$$

Where $f(x)$ corresponds to a series of conditional checks based on the features of the input data.

The tests utilized optimal PSD features extracted from the recorded acoustic data. To ensure the reliability of the outcomes, we performed both training and cross-validation by splitting the dataset into five parts. Figure 3 shows the training and testing process in one iteration. SMOTE is applied only in the training phase. The selection of the best features is also done only in the training stage. The trained model is tested with one type of PSD feature. The type of feature used is obtained from information during training. It is ensured that the data used in the testing phase differs from that used during training. The outcome is determined by assessing the model's accuracy during testing. The highest-performing PSD parameters for each class are stochastically partitioned into two distinct groups: a model-training cohort and an evaluation cohort. Predictive reliability is assessed using criteria such as F-score (a measure of precision and sensitivity), classification accuracy, actual negative rate (specificity), detection rate (recall), and overall correctness.

Calculation of Model Performance

Assessment is conducted within the supervised learning framework [30]. The confusion matrix structure assigns rows to represent the correct class labels and columns to denote the model-predicted classifications.

To refine error analysis, especially when distinct groups are mapped to the same categories, it is suggested to invert rows and columns in confusion matrices, placing predicted outcomes in rows and actual labels in columns. This provides a clearer view of misclassification patterns and enables more targeted improvements in model performance.

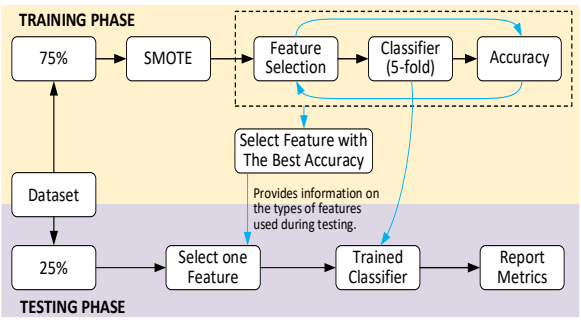

Figure 3. One iteration overview of the training and testing process.

Each cell in the matrix captures a specific pairing of real-world and algorithmic classifications. Correct identifications (TP) arise when predictions align with positive ground truths. Correct rejections (TN) occur when both actual and predicted labels are negative. Misclassifications include false alarms (FP), in which negatives are incorrectly flagged as positives, and misses (FN), in which positives are incorrectly dismissed as negatives.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

Evaluating model performance metrics: (15) can assess the accuracy of the model's predictions. Equation 5 is used to assess the model's performance, focusing on recall (sensitivity).

$$Sensitivity = \frac{TP}{TP + FN} \quad (16)$$

To compute how precise the model is in pinpointing the exact nature of the machine damage, as shown in (16).

$$Precision = \frac{TP}{TP + FP} \quad (17)$$

Equation 17 is applied to calculate the specificity parameter.

$$Specificity = \frac{TN}{TN + FP} \quad (18)$$

The F-score parameter is computed using (18).

$$\text{F-}score = \frac{2TP}{2TP + FP + FN} \quad (19)$$

RESULTS

Preprocessing and Feature Extraction Results

Before beginning feature extraction, it is important to preprocess and examine the data to ensure smooth implementation of the classification process. Figure 4 depicts machine sounds in the temporal domain. We randomly chose three machine noise samples from each category recorded in the same session. In the time domain, signals show similar magnitudes and distributions. Therefore, it is crucial to identify unique features to highlight pattern differences across various classes. To streamline computational demands, dimensionality reduction is achieved through feature selection. Figure 5 demonstrates the application of the Welch algorithm for spectral estimation, enabling bidirectional conversion between the temporal waveforms and frequency-domain representations of audio signals. This method optimizes the analysis of signal energy distribution across spectral bands.

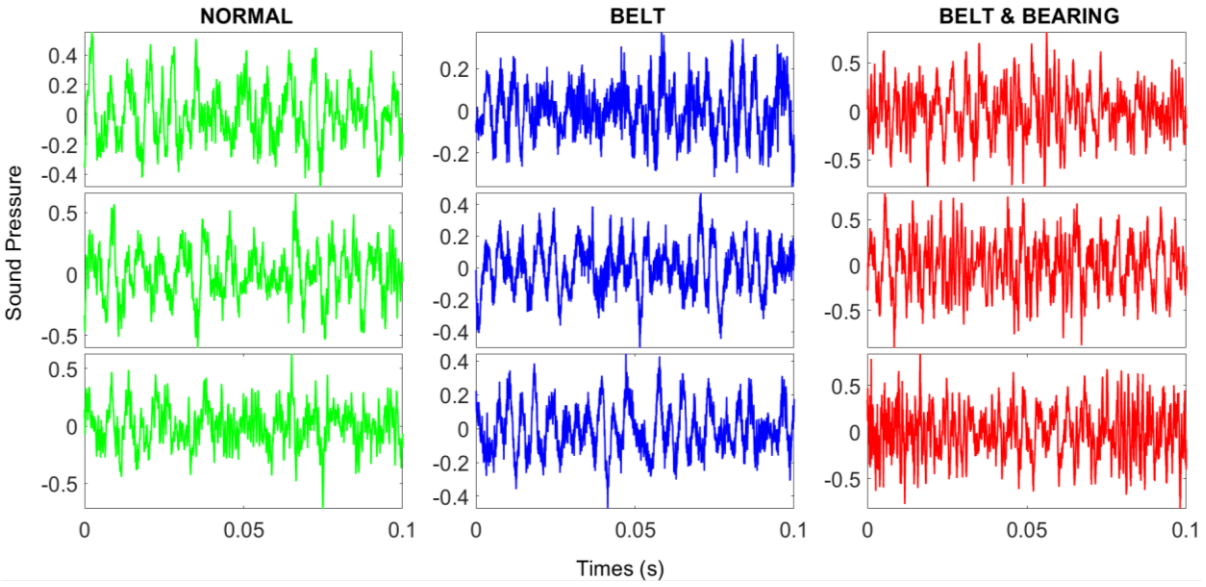

Figure 4. Machine audio in the time domain

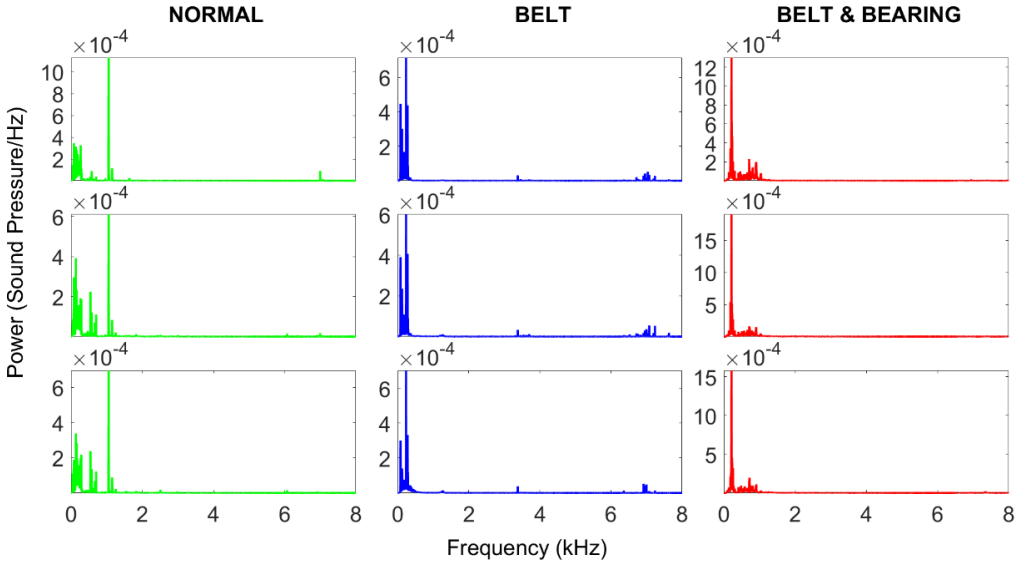

Figure 5. Power spectral density

Each fault state exhibits unique PSD signatures. Broadly, dominant energy concentration typically spans the 0–1 kHz spectral band. A loose van belt condition manifests a marked amplification of spectral energy below 1 kHz and near 7 kHz. Conversely, combined belt instability and bearing degradation reveal intensified energy within the 0–1 kHz range, with a distinct maximum near 0.3 kHz. This contrast highlights how mechanical faults alter frequency-specific energy distributions. Using PSD can help identify unique patterns specific to each category. Since the PSD dataset contains a large number of series, many of which are not helpful for classification, we used seven statistical equations as described in the classification chapter. To isolate the most relevant PSD features. This approach simplifies the workflow and boosts classification effectiveness.

Feature Selection Results

In machine learning and statistics, feature selection is the process of identifying and retaining only the most important features from a larger dataset. We calculate accuracy to determine seven essential parameters that improve the model's precision in classifying machine audio. Consequently, we identified the paramount characteristic of the categorization process. This pattern suggests that the classification model will have difficulty precisely discriminating between classes based on the given attributes. The first feature, along with any features derived from the mean of PSD, exhibits the highest accuracy across several classifiers, as illustrated in Figure 6. Feature 1 is the principal attribute used for categorization. The results demonstrate that the average PSD type feature exerts the most significant influence in differentiating among the various class types.

Based on the feature selection results, we use the average PSD feature across all categories during testing. Using fewer features in machine learning has several significant advantages. Models are more straightforward to understand, which improves interpretability. Models with fewer features tend to be faster to train and run, as they require fewer calculations and computational resources. Using many features in machine learning or deep learning classification can cause several problems. One of the main issues is overfitting, where the model becomes overly tailored to the training data, resulting in high accuracy on the training data but poor performance on new, unseen test data. In addition, the curse of dimensionality phenomenon arises: the more features, the sparser the data becomes, making it difficult for the model to find clear patterns. On the other hand, multicollinearity among correlated features can lead to redundancy in the information the model learns, distorting coefficient estimates and making interpretation difficult.

Classifier Performance

This is due to the difference in the number of remaining samples. We employ the SMOTE algorithm to impute missing samples. The average performance metrics shown in Figure 7 are used to evaluate the classification model's efficacy. Removing all attributes except the average PSD does not diminish the accuracy of each classifier model.

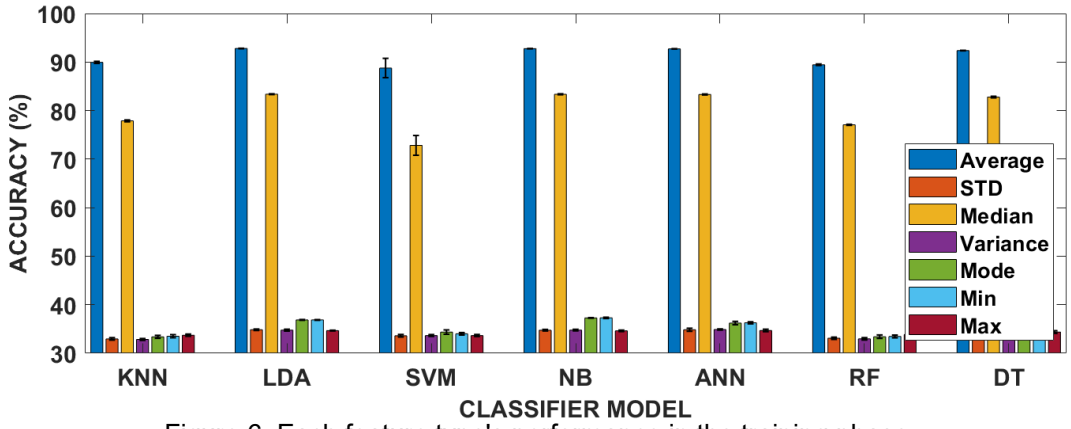

Figure 6. Each feature type's performance in the training phase

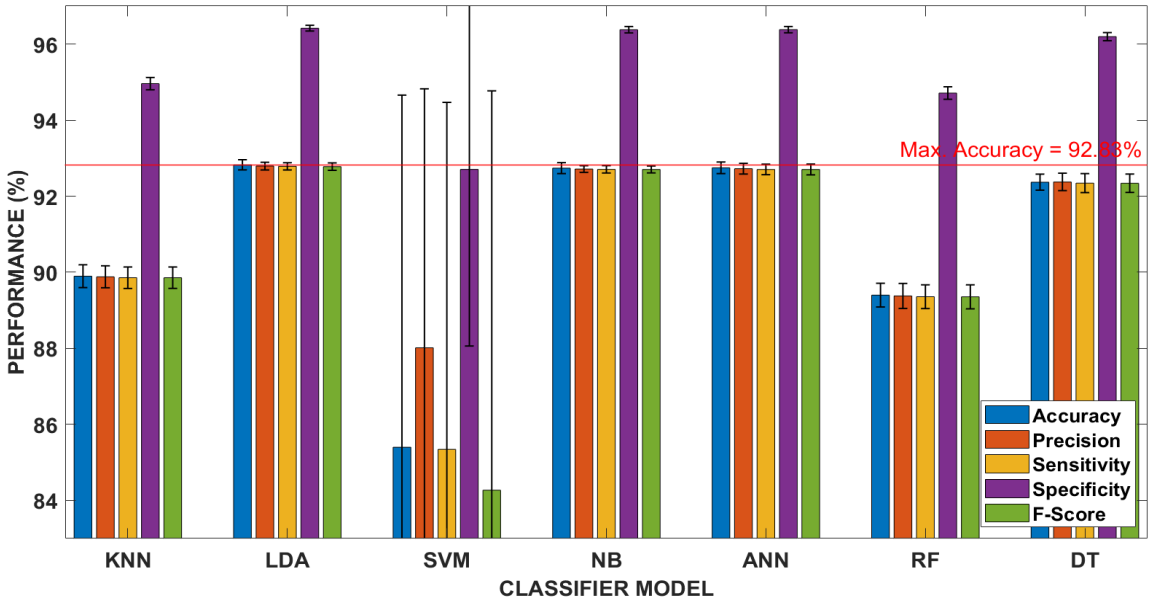

Figure 7. Each classifier model's performance in the testing phase

Misclassification most commonly occurs within the regular class. The performance comparison of seven classifier models, KNN, LDA, SVM, NB, ANN, RF, and DT, across five evaluation metrics: accuracy, precision, sensitivity, specificity, and F-score. Among these models, LDA achieved the highest accuracy at 92.83%, followed closely by NB at 92.74%, ANN at 92.75%, and DT at 92.34%. These four models outperform KNN (89.90%) and RF (89.40%), which showed relatively lower accuracy. SVM had the lowest accuracy of 85.40%, indicating weaker overall performance. Notably, the specificity of most models, especially LDA, NB, ANN, and DT, is notably high, exceeding 95%, suggesting strong performance in distinguishing between true negatives. While precision and sensitivity showed moderate variation, the F-score generally aligned well with accuracy, indicating balanced performance across the models.

However, LDA is very efficient in accurately detecting each machine condition with just one feature. The model achieves the highest accuracy across classes when classifying average PSD features. The average PSD precision was 92.79%, sensitivity was 92.79%, specificity was 96.42%, and the F-score was 92.78%. The results substantiate the effectiveness of our proposed methodology for detecting different types of machine failures by analyzing sound data.

DISCUSSION

The results presented in this study provide a compelling narrative of the effectiveness of the proposed approach for classifying machine faults from audio data. Figure 5 highlights the distinctive differences in the PSDs of machine sounds across various fault classes, providing a basis for robust feature extraction. Figure 6 clearly demonstrates that the average PSD consistently achieves the highest classification accuracy, validating its

strong discriminatory power. Furthermore, Figure 7 shows that using only the average PSD feature does not compromise classifier performance, particularly with the LDA model, which achieves an accuracy of 92.83%.

The logical progression from data inspection to feature extraction and selection underscores the robustness of the methodology. Initially, the raw time-domain signals exhibit similar magnitudes and distributions as shown in Figure 4, suggesting minimal distinguishing information. The transformation to the frequency domain reveals class-dependent patterns that are more pronounced. Recognizing the computational and interpretive challenges of high-dimensional data, we employed seven statistical equations, as written in the classification section, to extract the most salient features, ultimately identifying the average PSD as the most informative.

This finding demonstrates a direct causal relationship to the success of the final model: reducing the dimensionality of the feature space by selecting the average PSD not only accelerates computation but also mitigates the risk of overfitting and redundancy. The exceptional performance of the LDA classifier, achieving above 90% in all performance metrics, underscores the effectiveness of this feature selection strategy. The superior performance of LDA can be attributed to several factors, particularly its ability to handle linearly separable data. By projecting the data onto a lower-dimensional space where class distributions are well-separated, LDA excels when the assumption of linearity holds, making it particularly well-suited to the reduced, more focused feature set after dimensionality reduction.

This outcome also addresses concerns such as the "curse of dimensionality" and multicollinearity, which often arise in high-dimensional feature sets. In contrast, SVM and Random Forests RF may not perform as well in this context due to their more complex, non-linear decision boundaries and higher sensitivity to noise and overfitting when the feature space is ample. While SVM and RF are robust classifiers, they may struggle with increased dimensionality and inter-feature correlations, leading to less stable performance than LDA in this particular setup. Therefore, the combination of reduced dimensionality and LDA's assumption of linear separability provides a more reliable and computationally efficient model.

Compared with previous studies, our findings align with and extend the existing knowledge base. Prior work, such as that by Li et al. [2] and Wang et al. [10], employed vibration- and audio-based features and applied them to deep learning models for machine fault diagnosis. These studies reported high accuracy rates, often exceeding 90%, but relied on multiple features and complex architectures. Our work demonstrates that by leveraging a single well-chosen feature and a simpler classifier, we can achieve high accuracy levels. However, it is important to note that the current method was not tested on data from the previously mentioned works, nor was their model evaluated on our dataset. As a result, a direct comparison between the two approaches using the same data is not available. This outcome suggests that simpler models can rival more complex ones when the most informative feature is used, reinforcing the notion that model complexity does not always equate to better performance.

Moreover, our results challenge the prevailing reliance on deep learning models for audio classification in machine condition monitoring. Unlike these approaches, our methodology is faster, more interpretable, and requires significantly fewer computational resources. This directly supports arguments in recent literature advocating for more resource-efficient models in industrial settings [31][32]. Our findings thus offer a practical alternative for scenarios with constrained resources or real-time monitoring needs.

To summarize, our work contributes to the field by demonstrating that a single feature, the average PSD, can achieve high classification accuracy, challenging the assumption that complex feature sets are always necessary. LDA can outperform more computationally intensive models, suggesting that simpler models can be more robust when the features are highly discriminative. Sound-based monitoring can be a viable alternative to vibration-based approaches, offering a low-cost, accessible solution for machine health monitoring. These insights strengthen the argument for lightweight, interpretable models in industrial diagnostics, especially in settings where simplicity, speed, and cost-effectiveness are critical.

CONCLUSION

Using a cellphone, we successfully detected damage to the fan booster machine by carefully analyzing captured audio signals. Each machine state had recordings that extended beyond 3 hours. The architecture of the fan booster defect identification system was built through sequential phases: acoustic signal acquisition, preprocessing (including noise filtering and amplitude normalization), Welch-algorithm-driven spectral feature extraction, and evaluative refinement of discriminative attributes

using validation metrics. This methodology prioritizes computational efficiency while isolating fault-specific spectral patterns and comparing the results of seven different classification models. Among these models, the LDA classifier demonstrated the best performance, achieving 92.83% accuracy in classifying the average PSD feature. Notably, the minimum performance across all evaluated parameters consistently remained above 92%, indicating the approach's strength and reliability.

ACKNOWLEDGMENT

The authors extend their sincere appreciation to Universitas Mercu Buana for the generous financial support provided under contract No. 02-5/794/B-SPK/III/2024. This support played a crucial role in enabling the successful completion of this research work.

REFERENCES

[1] S. Gawde, S. Patil, S. Kumar, P. Kamat, K. Kotecha, and S. Alfarhood, “Explainable Predictive Maintenance of Rotating Machines Using LIME, SHAP, PDP, ICE,” *IEEE Access*, vol. 12, pp. 29345–29361, 2024, doi: 10.1109/ACCESS.2024.3367110.

[2] X. Li, W. Zhang, Q. Ding, and J. Q. Sun, “Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation,” *J Intell Manuf*, vol. 31, no. 2, pp. 433–452, Feb. 2020, doi: 10.1007/S10845-018-1456-1.

[3] D. Romahadi, A. A. Luthfie, W. Suprihatiningsih, and H. Xiong, “Designing expert system for centrifugal using vibration signal and Bayesian Networks,” *Int J Adv Sci Eng Inf Technol*, vol. 12, no. 1, pp. 23–31, 2022, doi: 10.18517/IJASEIT.12.1.12448.

[4] L. He, C. Yi, Q. Zhou, and J. H. Lin, “Fast Convolutional Sparse Dictionary Learning Based on LocOMP and Its Application to Bearing Fault Detection,” *IEEE Trans Instrum Meas*, vol. 71, 2022, doi: 10.1109/TIM.2022.3193962.

[5] J. Supriyono, I. Mukhlash, M. Iqbal, and D. A. Asfani, “TE-LSTM: winding temperature prediction for induction motors in the oil and gas industry,” *SINERGI*, vol. 29, no. 3, pp. 819–832, Sep. 2025, doi: 10.22441/sinergi.2025.3.022.

[6] J. Pacheco-Chérrez, J. A. Fortoul-Díaz, F. Cortés-Santacruz, L. María Aloso-Valerdi, and D. I. Ibarra-Zarate, “Bearing fault detection with vibration and acoustic signals: Comparison among different machine leaning classification methods,” *Eng Fail Anal*, vol. 139, p. 106515, Sep. 2022, doi: 10.1016/J.ENGFAILANAL.2022.106515.

[7] Y. Ren, Z. Zhou, W. Ren, B. Luo, and Z. Yue, “Impact damage and bearing capacity of CFRP laminates with pretensioned tooth joints and graded thickness,” *Results in Engineering*, vol. 28, p. 108095, Dec. 2025, doi: 10.1016/J.RINENG.2025.108095.

[8] Z. Lou *et al.*, “Formation of Corrugated Damage on Bearing Race under Different AC Shaft Voltages,” *Materials*, vol. 17, no. 4, Feb. 2024, doi: 10.3390/MA17040859/S1.

[9] S. N. Chegini, M. J. H. Manjili, and A. Bagheri, “New fault diagnosis approaches for detecting the bearing slight degradation,” *Meccanica*, vol. 55, no. 1, pp. 261–286, Jan. 2020, doi: 10.1007/s11012-019-01116-x.

[10] Y. S. Wang, N. N. Liu, H. Guo, and X. L. Wang, "An engine-fault-diagnosis system based on sound intensity analysis and wavelet packet preprocessing neural network," *Eng Appl Artif Intell*, vol. 94, p. 103765, Sep. 2020, doi: 10.1016/j.engappai.2020.103765.

[11] J. Park, S. Kim, J. H. Choi, and S. H. Lee, “Frequency energy shift method for bearing fault prognosis using microphone sensor,” *Mech Syst Signal Process*, vol. 147, p. 107068, Jan. 2021, doi: 10.1016/j.ymssp.2020.107068.

[12] P. Sharma, S. Chandra Rana, and R. Nath Barman, “Experimental study on fault analysis of a needle roller bearing with wear evaluation,” *Mater Today Proc*, vol. 24, pp. 567–575, Jan. 2020, doi: 10.1016/j.matpr.2020.04.310.

[13] N. Indah, J. Yunas, A. A. Hamzah, R. Latif, D. S. Khaerudini, and Subekti, “Analyzing the Vibration and Mechanical Properties of the MEMS Membrane with Planar Spring Arms Design for Energy Harvesting,” *AIP Conf Proc*, vol. 3115, no. 1, Sep. 2024, doi: 10.1063/5.0207450/3313646.

[14] Y. Wu *et al.*, “The Fault Diagnosis of Rolling Bearings Based on FFT-SE-TCN-SVM,” *Actuators 2025, Vol. 14, Page 152*, vol. 14, no. 3, p. 152, Mar. 2025, doi: 10.3390/ACT14030152.

[15] C. Pichler *et al.*, “Acoustic Condition Monitoring: Signal Analysis for Large Machinery Halls,” *Conference Record - IEEE Instrumentation and Measurement Technology Conference*, 2022, doi: 10.1109/I2MTC48687.2022.9806680.

[16] T. Chen, L. Guo, H. Gao, T. Feng, and Y. Yu, “Clustering Weighted Envelope Spectrum for Rolling Bearing Fault Diagnosis,” *IEEE Transactions on Automation Science and*

Engineering, 2024, doi: 10.1109/TASE.2024.3403665.
[17] M. Kuncan, "An Intelligent Approach for Bearing Fault Diagnosis: Combination of 1D-LBP and GRA," *IEEE Access*, vol. 8, pp. 137517–137529, Jan. 2020, doi: 10.1109/ACCESS.2020.3011980.
[18] N. K. Hosseini, H. Toshani, S. Abdi, and S. Sharifzadeh, "Enhanced Bearing Fault Detection in Induction Motors Using Projection-Based SVM," *IEEE Trans Ind Appl*, 2025, doi: 10.1109/TIA.2025.3536425.
[19] T. Chen and D. Yu, "A Novel Method for Enhanced Demodulation of Bearing Fault Signals Based on Acoustic Metamaterials," *IEEE Trans Industr Inform*, vol. 18, no. 10, pp. 6857–6864, Oct. 2022, doi: 10.1109/TII.2022.3143161.
[20] The MathWorks Inc., "MATLAB Version: 23.2.0.2409890 (R2023b)," 2023, *The MathWorks Inc.*
[21] A. S. Siva, S. G. Rameshkumar, and K. Dhayalini, "Supraharmonic Analysis by Welch's-Power Spectral Density Estimation," *2024 3rd International Conference on Power, Control and Computing Technologies, ICPC2T 2024*, pp. 357–362, 2024, doi: 10.1109/ICPC2T60072.2024.10474764.
[22] C. Song, T. Zhao, L. Xu, and X. Huang, "Probabilistic prediction of uniaxial compressive strength for rocks from sparse data using Bayesian Gaussian process regression with Synthetic Minority Oversampling Technique (SMOTE)," *Comput Geotech*, vol. 165, p. 105850, Jan. 2024, doi: 10.1016/J.COMPGEO.2023.105850.
[23] R. K. Halder, M. N. Uddin, M. A. Uddin, S. Aryal, and A. Khraisat, "Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications," *Journal of Big Data*, vol. 11, no. 1, pp. 1–55, Aug. 2024, doi: 10.1186/S40537-024-00973-Y.
[24] L. Qu and Y. Pei, "A Comprehensive Review on Discriminant Analysis for Addressing Challenges of Class-Level Limitations, Small Sample Size, and Robustness," *Processes 2024, Vol. 12, Page 1382*, vol. 12, no. 7, p. 1382, Jul. 2024, doi: 10.3390/PR12071382.
[25] J. Rybak and W.-X. Zhou, "On Inference for the Support Vector Machine," *Journal of Machine Learning Research*, vol. 26, pp. 1–54, 2025, Accessed: Aug. 26, 2025. [Online]. Available: http://jmlr.org/papers/v26/23-1581.html.
[26] H. Chen, S. Hu, R. Hua, and X. Zhao, "Improved naive Bayes classification algorithm for traffic risk management," *EURASIP J Adv Signal Process*, vol. 2021, no. 1, pp. 1–12, Dec. 2021, doi: 10.1186/S13634-021-00742-6/TABLES/5.
[27] M. Skrbek, P. Kubalik, M. Kohlik, J. Borecky, and R. Hulle, "Evaluation of the Medium-sized Neural Network using Approximative Computations on Zynq FPGA," *12th Mediterranean Conference on Embedded Computing, MECO 2023*, 2023, doi: 10.1109/MECO58584.2023.10155065.
[28] Y. Yuan, K. Wang, D. Duives, W. Daamen, and S. P. Hoogendoorn, "Machine learning-based bicycle delay estimation at signalized intersections using sparse GPS data and traffic control signals - A Dutch case study using random forest algorithm," *Artificial Intelligence for Transportation*, vol. 3–4, p. 100037, Nov. 2025, doi: 10.1016/J.AIT.2025.100037.
[29] Z. R. S. Elsi, A. A. Supli, Jimmie, M. G. Al-Faris, and D. A. Rapel, "Optimizing intrusion detection with data balancing and feature selection techniques," *SINERGI*, vol. 29, no. 3, pp. 779–792, Sep. 2025, doi: 10.22441/sinergi.2025.3.019.
[30] M. Hasnain, M. F. Pasha, I. Ghani, M. Imran, M. Y. Alzahrani, and R. Budiarto, "Evaluating Trust Prediction and Confusion Matrix Measures for Web Services Ranking," *IEEE Access*, vol. 8, pp. 90847–90861, 2020, doi: 10.1109/ACCESS.2020.2994222.
[31] X. Liu, D. Pei, G. Lodewijks, Z. Zhao, J. M.-A. P. Technology, and undefined 2020, "Acoustic signal based fault detection on belt conveyor idlers using machine learning," *Elsevier*, Accessed: Nov. 20, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0921883120301898
[32] B. Chen, L. Xie, Y. Li, and B. Gao, "Acoustical damage detection of wind turbine yaw system using Bayesian network," *Renew energy*, vol. 160, pp. 1364–1372, Nov. 2020, doi: 10.1016/j.renene.2020.07.062.