

Supervised Classification of Industrial Fan Sound Anomalies Using Neural Networks and Engineered Acoustic Features

Angelina Pramana Thenata¹, Ranny², Bhustomy Hakim³, Fergie Joanda Kaunang^{4,*}

^{1,2,3,4}Faculty of Design and Technology, Universitas Bunda Mulia
Jl. Lodan Raya No.2, Ancol, Jakarta, Indonesia

¹athenata@bundamulia.ac.id

²ranny.fransiska@gmail.com

³bhakim@bundamulia.ac.id

*Correspondence: fkaunang@bundamulia.ac.id

Abstract— This study uses a supervised learning approach based on neural networks for anomaly detection in industrial fan systems. Using a subset of the FAN data from the MIMII (malfunctioning industrial machine investigation and inspection) dataset with 530 labelled recordings (383 normal and 147 abnormal), this study extracts acoustic features including mel-frequency cepstral coefficients (MFCC), spectral descriptors (centroid, roll off), and temporal measures (zero-crossing rate, autocorrelation). Univariate statistical tests reveal that several MFCC coefficients and time-domain features differ significantly between classes ($p < 0.05$). A feed-forward neural network model with two hidden layers of 64 units (ReLU activation) and dropout regularisation was trained using stratified cross-validation with 5-fold, resulting in an average F1 score of 89.9%. The use of several threshold values ($\tau \in \{0.3-0.7\}$) confirmed the robustness of the model, as seen in the test data results with the selected threshold value of $\tau = 0.5$, which achieved a precision of 100%, recall = 93.10%, F1 = 96.43%, and accuracy = 98.11% (identical results were obtained at $\tau = 0.6-0.7$; while $\tau = 0.3$ provided higher recall). The model also produced an AUC-ROC value of 0.9978, which is close to ideal and demonstrates excellent cross-threshold discrimination. These findings demonstrate that combining interpretable acoustic features with a compact neural classifier enables accurate non-invasive anomaly detection for Industry 4.0 applications with minimal hardware requirements.

Keywords— industrial fan, anomaly detection, acoustic features, neural network, supervised learning

Abstract— Penelitian ini menggunakan pendekatan supervised learning berbasis jaringan saraf untuk deteksi anomali pada sistem kipas industri. Dengan subset data FAN dari MIMII (malfunctioning industrial machine investigation and inspection) dataset dengan 530 rekaman berlabel (383 normal dan 147 abnormal), penelitian ini mengekstraksi fitur akustik yang meliputi mel-frequency cepstral coefficients (MFCC), spectral descriptor (centroid, roll off), serta temporal measures (zero-crossing rate, autocorrelation). Uji statistik univariat menunjukkan sejumlah koefisien MFCC dan fitur domain waktu berbeda signifikan antar kelas ($p < 0,05$). Model jaringan saraf feed-forward dengan dua lapisan tersembunyi berukuran 64 unit (aktivasi ReLU) dan regularisasi dropout dilatih menggunakan stratified cross validation dengan 5-fold sehingga menghasilkan nilai F1 rata-rata sebesar 89,9%. Penggunaan beberapa nilai ambang ($\tau \in \{0,3-0,7\}$) menegaskan kekokohan model yang terlihat pada hasil data uji dengan nilai ambang terpilih adalah $\tau = 0,5$ yang mencapai precision sebesar 100%, recall = 93,10%, F1 = 96,43%, dan akurasi = 98,11% (hasil identik diperoleh pada $\tau = 0,6-0,7$; sementara $\tau = 0,3$ memberikan recall lebih tinggi). Model juga menghasilkan nilai AUC-ROC sebesar 0,9978 yang mendekati ideal dan menunjukkan daya diskriminasi lintas-ambang yang sangat baik. Temuan ini memperlihatkan bahwa penggabungan fitur akustik yang dapat diinterpretasikan dengan pengklasifikasi saraf yang ringkas

memungkinkan deteksi anomali non-invasif yang akurat untuk penerapan Industri 4.0 dengan kebutuhan perangkat keras minimal.

Kata Kunci — *kipas mesin industri, deteksi anomali suara, fitur akustik, jaringan saraf tiruan, supervised learning*

I. INTRODUCTION

Industrial fans are essential components in numerous manufacturing processes and industrial facilities, serving critical functions such as ventilation, cooling, airflow regulation, and environmental temperature control. Any operational malfunction in these fans can trigger cascading failures—resulting in production inefficiencies, increased maintenance costs, potential equipment damage, and safety hazards [1]. Consequently, reliable condition monitoring systems and proactive maintenance strategies are crucial for ensuring safe, efficient, and uninterrupted industrial operations. Traditionally, the health of industrial machinery is monitored through manual inspection routines or by deploying vibration and temperature sensors. Skilled technicians rely on their sensory perceptions—such as vision, hearing, smell, and touch—combined with professional expertise to identify irregularities. While these conventional approaches offer adaptability and immediate interpretation, they are inherently limited by ambiguous diagnostic standards, inconsistencies in human judgment, and challenges in scaling workforce availability to balance operational demands and costs [2], [3].

With the increasing adoption of artificial intelligence (AI), especially machine learning (ML), data-driven methods have emerged as powerful tools to automate fault detection and classification. In parallel with these advancements, the growing adoption of artificial intelligence (AI), particularly in the domain of machine learning (ML), has accelerated the transition from manual inspection to automated anomaly detection and classification systems. Machine learning techniques have been widely applied across diverse domains, including web-based applications for recognizing handwritten Javanese script using convolutional neural networks (CNNs) [4], music genre clustering through unsupervised learning [5], and financial forecasting using hybrid models such as temporal convolutional networks and generative adversarial networks (TCN-GAN) [6]. These innovations demonstrate the flexibility and robustness of ML in capturing complex patterns across various types of data.

In the context of industrial maintenance, ML algorithms are capable of learning normal operational behaviour from historical sensor data and identifying subtle deviations that may indicate potential faults. Specifically, in acoustic monitoring applications, machine-generated sound signals can be digitally analyzed to capture imperceptible variations that may escape human detection. This capability allows ML-based diagnostic systems to achieve consistent, repeatable, and scalable monitoring performance that is not limited by human subjectivity [7]. These approaches learn patterns of normal machine behavior from historical data and can detect subtle deviations indicative of potential anomalies. Among the various sensing modalities, acoustic-based monitoring stands out due to its non-invasive nature, cost-effectiveness, and capacity to capture early mechanical faults through sound signatures before they manifest physically [8]. Machine learning models are well-suited for interpreting audio data by extracting representative features such as mel-frequency cepstral coefficients (MFCCs), spectral centroid, spectral roll-off, and zero-crossing rate. These features encapsulate important temporal and frequency domain information from audio signals, enabling the training of classifiers that can discriminate between normal and faulty machine states [9], [10].

While classical machine learning algorithms such as support vector machines (SVMs) and random forests have demonstrated success in this domain [11], [12], deep learning models like convolutional neural networks (CNNs) and autoencoders have shown superior performance due to their ability to learn complex,

nonlinear patterns from high-dimensional data. Recent studies have explored audio-based anomaly detection across a range of industrial machines. For example, [13] introduces a novel method for sound-based anomaly detection aimed at overcoming the limitations of traditional, costly sensor-based systems and the challenge of industrial noise. Likewise, [8] proposes an intelligent fault diagnosis system for hydraulic piston pumps using a novel PSO-LeNet model. Applied to acoustic signals, the model successfully identifies five common pump faults. Comparative analysis against several deeper CNNs, including AlexNet and VGG models, showed that the PSO-LeNet achieved superior stability and the highest accuracy, demonstrating its excellent overall performance. Although these unsupervised approaches are advantageous in settings with limited labeled data, they frequently suffer from high false-positive rates due to their assumption that all deviations from the learned normal pattern signify anomalies. In contrast, supervised learning approaches are capable of more precise and robust classification when labeled data for both normal and anomalous classes is available [14].

Despite the abundance of studies focusing on pumps [15], [16], [17], valves [18], [19], and motors [20], [21] limited research has specifically addressed anomaly classification in industrial fan systems, which often have unique acoustic characteristics and operate in environments with high background noise. Additionally, fan faults such as bearing wear, imbalance, and misalignment, can manifest subtly in audio signals, posing a challenge for generalized models. To address this gap, the present study proposes a supervised neural network model for the binary classification of industrial fan sounds using engineered acoustic features. This approach integrates domain expertise in audio signal processing with the learning capabilities of neural networks to effectively distinguish between normal and abnormal fan operating conditions. The study leverages the publicly available malfunctioning industrial machine investigation and inspection (MIMII) dataset [22], which contains labeled recordings of machine sounds under real-world factory conditions, including controlled faults and normal operations.

An accurate and lightweight acoustic classifier has significant implications for industry 4.0 and predictive maintenance frameworks, enabling scalable, real-time monitoring solutions without the need for expensive or invasive sensor systems. The model presented in this research aims to contribute toward the development of intelligent diagnostic systems suitable for integration into modern industrial infrastructure. The remainder of this paper is organized as follows: Section II describes the dataset, feature extraction, and proposed methodology; Section III presents the experimental setup and results; and Section IV concludes the study and outlines future work directions.

II. METHODOLOGY

Fig. 1 shows the flow of this study which employed a supervised learning approach to classify industrial fan sound anomalies using a neural network model. The methodology consist of six stages: data collection, data preprocessing, exploratory data analysis, feature extraction, model development, and evaluation.

1) *Data Collection*: This research begins with the acquisition of industrial machine sound data, sourced from the MIMII (malfunctioning industrial machine investigation and inspection) dataset [22]. This dataset specifically contains audio recordings from industrial fan machines under two primary operating conditions: normal and anomalous, where the latter is caused by mechanical faults or malfunctions. In this study, a total of 530 audio samples were utilized, consisting of 383 normal and 147 abnormal recordings.

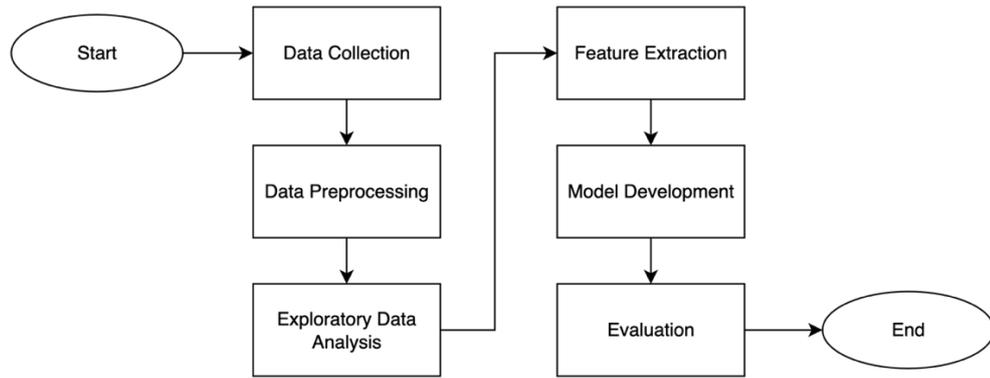


Fig. 1 Research flow

2) *Data Preprocessing*: Audio samples were sourced from the MIMII dataset, comprising 530 recordings (383 normal, 147 abnormal) in .wav format. Each file was loaded using the Librosa library, extracting the time-series waveform and sampling rate. Binary labels (0 for normal, 1 for abnormal) were assigned to facilitate supervised learning.

3) *Exploratory Data Analysis (EDA)*: An exploratory data analysis (EDA) was performed to characterize normal vs. abnormal fan acoustics and assess the discriminative value of the extracted descriptors. Distributional analyses (histograms and kernel density estimates) revealed clear shifts for most mel-frequency cepstral coefficients (MFCCs), spectral centroid, spectral roll-off, and zero-crossing rate between classes, indicating potential separability. A correlation heatmap is used to show strong intra-correlation among features. Descriptive statistics (mean and standard deviation) were computed per class, and independent two-sample t-tests were applied feature-wise to evaluate the statistical significance of mean differences.

4) *Feature Extraction*: To convert raw audio signals into machine-interpretable representations, we extracted a comprehensive set of engineered acoustic features. mel-frequency cepstral coefficients (MFCCs) were computed as 13-dimensional vectors capturing spectral envelope characteristics, with mean values calculated across time frames to represent timbral properties. Spectral features included three key descriptors: spectral centroid (quantifying signal brightness), spectral roll-off (measuring frequency spread), and spectral contrast (encoding peak-valley dynamics in the frequency spectrum). Temporal characteristics were represented through zero-crossing rate (ZCR), which indicates signal noisiness, and autocorrelation for periodicity analysis. All features were normalized and concatenated into a unified feature vector for each audio sample, forming the input space for subsequent classification tasks. This multi-domain feature ensemble effectively encapsulated both spectral and temporal patterns indicative of fan operational states.

5) *Model Development*: The classification model was implemented as a feedforward neural network using TensorFlow Keras. The architecture consisted of an input layer matching the feature dimensions, followed by two dense hidden layers with 64 neurons each, employing ReLU activation functions to introduce nonlinearity. Dropout layers ($p = 0.5$) were incorporated between hidden layers to prevent overfitting through regularization. The output layer utilized a single neuron with sigmoid activation for binary classification. The model was trained using the Adam optimizer with binary cross-entropy loss function, selected for its effectiveness in probabilistic classification tasks. To ensure robust evaluation, we employed stratified 5-fold cross-validation, maintaining class distribution across folds. Training proceeded for 100 epochs with a batch size of 32, with early stopping monitored to prevent overfitting while ensuring convergence.

6) *Model Evaluation*: The proposed classification model was evaluated through a multi-stage process to ensure reliable and unbiased performance assessment. Initially, stratified 5-fold cross-validation was employed to address class imbalance and evaluate generalization. In each fold, the model was trained, validated, and its performance measured using the F1 score, a balanced metric well-suited for imbalanced binary classification. The average F1 score across folds provided a robust estimate of the model's consistency. Subsequently, final testing was performed on a separate, unseen test set to assess real-world performance. Probabilistic outputs were evaluated across multiple thresholds (0.3 to 0.7), and metrics including Accuracy, Precision, Recall, and F1 Score were computed. This multi-threshold analysis enabled examination of the model's behaviour under varying decision boundaries, crucial for applications where false positives or false negatives carry different costs. Finally, the area under the receiver operating characteristic curve (AUC-ROC) was calculated to measure threshold-independent discriminatory ability. A high AUC-ROC value confirmed the model's strong capability in distinguishing between normal and abnormal instances across all decision thresholds.

III. RESULT AND DISCUSSION

A. Feature Analysis

The evaluation of the proposed classification model was conducted in a multi-stage process to ensure a thorough and unbiased assessment of its performance. Initially, the model was validated using stratified 5-fold cross-validation. To gain deeper insight into the discriminative potential of extracted acoustic features, a series of visualizations were generated to compare the distributions between normal and abnormal machine sound samples, as illustrated in Fig. 2. The figure presents histograms and kernel density estimations for 13 mel-frequency cepstral coefficients (MFCCs), along with key spectral and temporal features including spectral centroid, spectral roll-off, spectral contrast, zero crossing rate, and autocorrela-

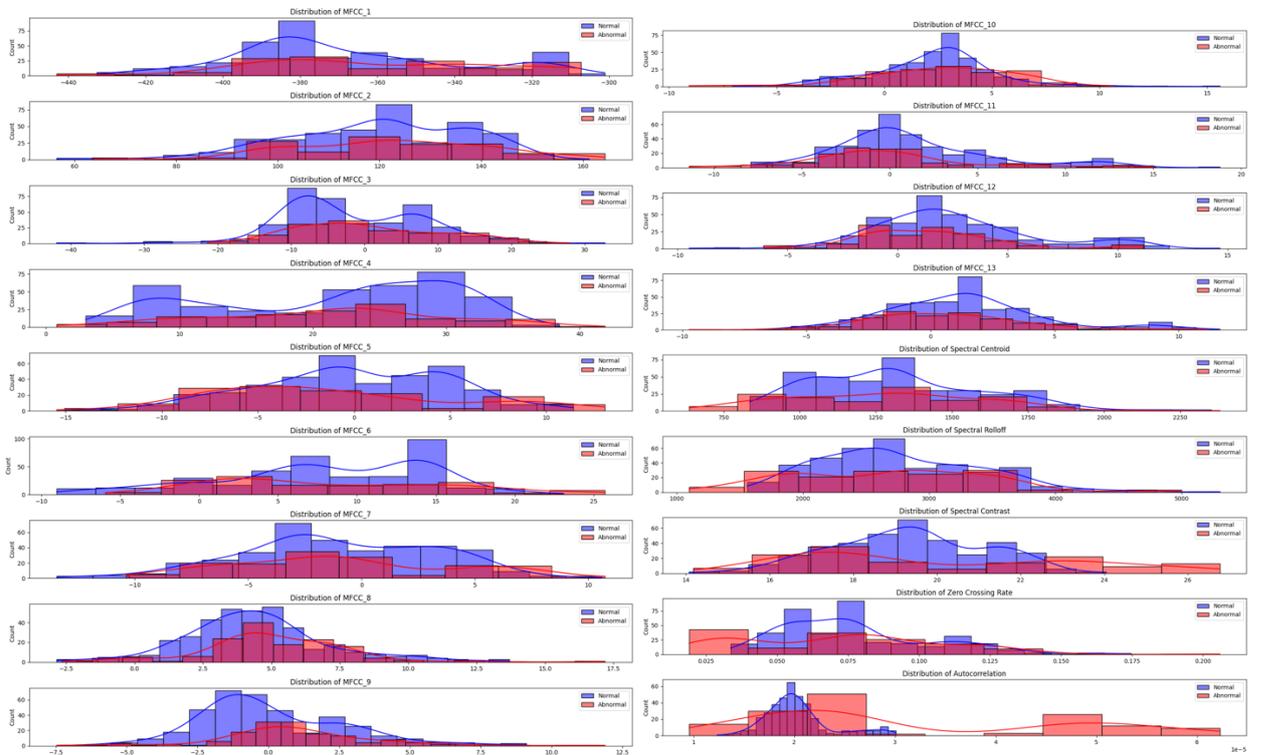


Fig. 2 Distribution plots of MFCCs and other acoustic features for normal (blue) and abnormal (red) classes

tion. Each feature was plotted for both the normal class (in blue) and the abnormal class (in red) to visually assess the separation between categories. The MFCC features, which are widely used to capture perceptual aspects of sound, demonstrate varying levels of separability. Specifically, MFCC_1 to MFCC_6 show clear differences in distribution shape and central tendency between normal and abnormal samples, suggesting strong discriminative power. In contrast, MFCC_11 to MFCC_13 exhibit considerable overlap, indicating limited contribution to classification performance. These results are consistent with statistical t -test outcomes, where only MFCCs with significant p -values ($p < 0.05$) were deemed informative.

Spectral centroid and spectral roll-off reveal substantial distributional shifts between the two classes. Abnormal sounds generally exhibit lower centroid and roll-off values, indicating energy concentration in lower frequency bands, which may result from mechanical degradation or frictional anomalies. Spectral contrast, however, presents overlapping distributions for both classes, aligning with its non-significant statistical difference and limited relevance for the classification task. The zero crossing rate demonstrates distinct behaviour for abnormal samples, showing a higher frequency of sign changes, which can be attributed to increased randomness or mechanical instability. Autocorrelation also displays a notable difference in distribution peaks, reflecting changes in signal periodicity often caused by anomalies.

To further understand the relationships among the extracted acoustic features, a Pearson correlation heatmap was constructed, as shown in Fig. 3. We quantify pairwise linear associations using the Pearson correlation coefficient r :

$$r(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

This analysis provides valuable insights into the redundancy, complementarity, and potential multicollinearity between features, which are critical considerations in the development of a robust classification model. The results indicate that the spectral centroid shows a strong positive correlation with MFCC_1 ($r = 0.81$), which is consistent with the interpretation of MFCC_1 as representing the overall energy or brightness of a sound signal, properties that are also encapsulated by the spectral centroid.

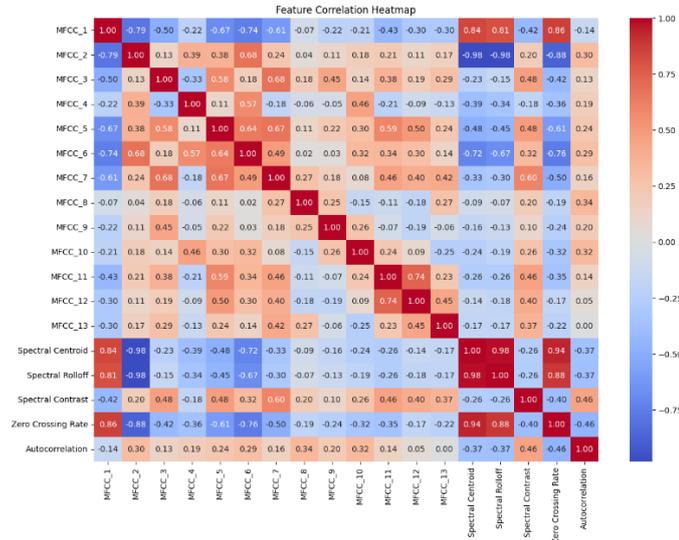


Fig. 3 Pearson correlation heatmap of extracted acoustic features. Blue indicates negative correlation, red indicates positive correlation, with values ranging from -1 to 1 .

Additionally, the spectral centroid demonstrates moderate positive correlations with MFCC_4 ($r = 0.58$), MFCC_5 ($r = 0.67$), MFCC_6 ($r = 0.64$), and MFCC_8 ($r = 0.57$), suggesting shared information content across these frequency-related features. In contrast, the spectral roll-off exhibits a strong negative correlation with both MFCC_1 ($r = -0.98$) and the spectral centroid itself ($r = -0.98$). This inverse relationship is expected, as signals with higher frequency content (i.e., higher centroid values) tend to have their spectral energy more evenly distributed, resulting in a higher roll-off threshold. Furthermore, spectral roll-off displays moderate negative correlations with other MFCCs, reinforcing its inverse association with frequency-distributed energy characteristics. The spectral contrast demonstrates generally weak correlations with most MFCCs and other spectral features, suggesting that it captures complementary information related to the variations between spectral peaks and valleys. Notably, it shows a moderate positive correlation with MFCC_4 ($r = 0.60$) and a moderate negative correlation with MFCC_1 ($r = -0.42$), highlighting its unique contribution to the feature space. Regarding temporal features, the zero crossing rate (ZCR) shows strong positive correlations with MFCC_1 ($r = 0.86$), spectral centroid ($r = 0.94$), and spectral roll-off ($r = 0.97$). These results are consistent with the idea that audio signals with greater high-frequency energy content tend to have more frequent zero crossings. Therefore, ZCR acts as an effective indicator of spectral sharpness or brightness.

The autocorrelation feature, which captures periodicity in the time domain, presents moderate negative correlations with MFCC_1 ($r = -0.42$), spectral centroid ($r = -0.37$), and spectral roll-off ($r = -0.46$). This suggests that signals with more harmonic or regular structures (i.e., higher autocorrelation values) are likely associated with lower-frequency energy distributions. Additionally, autocorrelation shows moderate positive correlations with MFCC_5 ($r = 0.41$) and MFCC_6 ($r = 0.40$), indicating potential relationships with mid-frequency cepstral dynamics.

The statistical analysis of extracted acoustic features, comprising mel-frequency cepstral coefficients (MFCCs), spectral, and temporal features, reveals meaningful distinctions between normal and abnormal sound classes as shown in Table I. These differences provide quantitative evidence that supports the suitability of these features for anomaly classification tasks. For the MFCC features, which characterize the spectral envelope of audio signals, significant variations in mean values were observed between the two classes. MFCC_1, which typically represents overall signal energy, shows a lower mean in normal samples (-369.33) compared to abnormal samples (-360.03), suggesting higher average energy or brightness in anomalous signals. Similarly, MFCC_2 and MFCC_3 display slightly elevated means in abnormal samples, indicating subtle shifts in spectral shape. The most pronounced difference occurs in MFCC_5, where the mean in the abnormal class (-1.31) is considerably lower than in the normal class (-0.04), with a notably higher standard deviation (6.22 vs. 4.99), indicating greater variability in anomalous signals.

MFCCs 6 through 10 also exhibit small yet consistent differences in means, generally trending higher for abnormal sounds, reinforcing the hypothesis of spectral distortion due to mechanical anomalies. MFCC_9 and MFCC_10, for instance, both show higher means in the abnormal class (1.04 vs. 0.07 and 2.61 vs. 2.18 , respectively), along with increased standard deviations, pointing to instability in the spectral structure under faulty conditions. Although MFCC_11 to MFCC_13 show more modest variations, their distributions still reflect subtle class-dependent differences. Spectral features further highlight differences in frequency characteristics. The spectral centroid, which indicates the center of mass of the frequency spectrum, is higher in normal sounds (1318.97 Hz) than in abnormal ones (1277.60 Hz), suggesting that normal sounds tend to be sharper or contain more high-frequency components. A similar pattern is observed in the spectral roll-off, which defines the frequency below which 85% of the signal energy is contained. Normal samples have a higher roll-off mean (2763.33 Hz) than abnormal samples (2656.77 Hz), supporting the notion of broader spectral content in normal operation. The spectral contrast, which measures the difference between spectral peaks and valleys, shows a slightly higher mean for abnormal sounds (19.98 vs.

19.31), along with a much larger standard deviation (3.36 vs. 1.82), indicating irregularities and greater variation in spectral dynamics under fault conditions.

Temporal features contribute valuable insights for anomaly classification. The zero crossing rate (ZCR) is slightly higher in normal sounds (0.08) compared to abnormal ones (0.07), with greater variability observed in the abnormal class. This suggests more consistent high-frequency content in normal signals and increased temporal irregularity in abnormal ones. While autocorrelation yields near-zero mean and standard deviation, statistical testing revealed it to be highly significant ($p \approx 3.95e-34$), highlighting its utility in capturing differences in signal periodicity. Independent t -tests confirmed that several MFCCs (MFCC_1, 3, 5, 8, 9, 11, 12, 13) significantly differ between normal and abnormal sounds ($p < 0.05$), indicating strong discriminative potential. In contrast, MFCC_2, 4, 6, 7, and 10 did not exhibit significant differences. Among spectral features, only spectral contrast showed statistical significance ($p = 0.00364$), while spectral centroid and roll-off did not, likely due to higher variance in the abnormal data. The model's performance was evaluated using stratified 5-fold cross-validation, yielding F1 scores between 0.8929 and 0.9123, with an average of 0.8989. This narrow range demonstrates the model's robustness and stability across data splits. An average F1 score near 0.90 confirms a well-balanced precision-recall trade-off, making the model reliable for real-world anomaly detection in industrial fan systems.

TABLE I
MEAN AND STANDARD DEVIATION OF EXTRACTED FEATURES FOR NORMAL AND ABNORMAL CLASSES

Feature	Normal Mean	Normal Std Dev	Abnormal Mean	Abnormal Std Dev
MFCC_1	-369.33	28.54	-360.03	31.69
MFCC_2	119.54	17.68	122.33	19.25
MFCC_3	-1.54	9.59	0.43	9.74
MFCC_4	21.48	8.96	21.00	8.75
MFCC_5	-0.04	4.99	-1.31	6.22
MFCC_6	8.07	6.57	8.30	7.53
MFCC_7	-0.99	4.27	-0.78	4.95
MFCC_8	4.45	2.63	5.07	2.59
MFCC_9	0.07	2.58	1.04	2.77
MFCC_10	2.18	2.85	2.61	3.37
MFCC_11	1.24	4.52	-0.09	4.75
MFCC_12	2.55	3.73	1.44	3.31
MFCC_13	1.28	3.14	0.56	3.11
Spectral centroid	1318.97	283.98	1277.60	325.52
Spectral roll-off	2763.33	637.96	2656.77	760.85
Spectral contrast	19.31	1.82	19.98	3.36
Zero crossing rate	0.08	0.03	0.07	0.04
Autocorrelation	0.00	0.00	0.00	0.00

Fig. 4 illustrates the training process of the proposed neural network model over 100 epochs, showing the progression of accuracy and loss. The left plot depicts the model's accuracy, which demonstrates a consistent upward trend, rising from approximately 0.64 to above 0.96. This steady increase indicates effective learning during the training phase, with the model progressively capturing discriminative patterns in the feature space. The right plot displays the corresponding training loss, which exhibits a sharp decline from an initial value of around 0.65 to below 0.1, stabilizing after approximately 80 epochs. The continual reduction in loss values, without abrupt oscillations or divergence, confirms that the learning rate and model complexity were appropriately configured, avoiding both underfitting and overfitting within the training data. The combination of increasing accuracy and decreasing loss suggests successful convergence of the training process. Moreover, the relatively smooth learning curves imply that the model generalizes well within the training set. To further validate generalizability, performance on a separate test set and cross-validation scores are reported in the following sections.

In binary classification for anomaly detection, the default probability cutoff ($\tau = 0.5$) may not be optimal. We therefore evaluate the classifier across multiple thresholds ($\tau \in \{0.3, 0.4, 0.5, 0.6, 0.7\}$) to characterize precision–recall trade-offs. As summarized in Table II, accuracy is consistently high and peaks at 98.11% for $\tau \geq 0.5$. At $\tau = 0.3$, the model attains higher recall = 96.55%—capturing more true anomalies—at the cost of lower precision = 93.33% (F1 = 94.92%, accuracy 97.17%), reflecting a small increase in false alarms. At $\tau = 0.4$, precision = 96.43% and recall = 93.10% (F1 = 94.74%, accuracy 97.17%), offering a slightly more conservative operating point. For $\tau \in \{0.5, 0.6, 0.7\}$, precision reaches 100% (no false positives

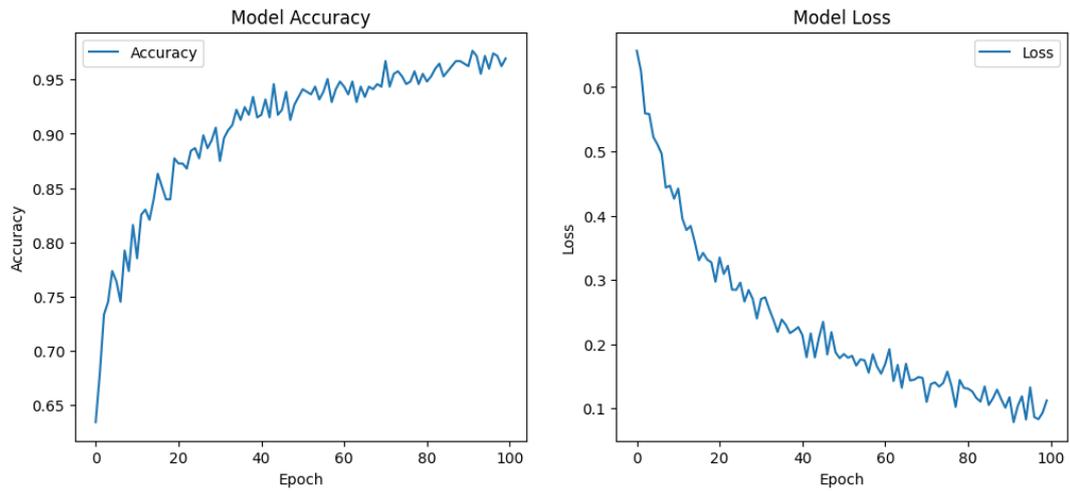


Fig. 4 Training accuracy and loss curves of the proposed neural network model over 100 epochs.

TABLE II
PERFORMANCE METRICS AT VARIOUS CLASSIFICATION THRESHOLDS

Threshold	Accuracy	Precision	Recall	F1 Score
0.3	97.17%	93.33%	96.55%	94.92%
0.4	97.17%	96.43%	93.10%	94.74%
0.5	98.11%	100%	93.10%	96.43%
0.6	98.11%	100%	93.10%	96.43%
0.7	98.11%	100%	93.10%	96.43%

in the test set) with recall = 93.10%, yielding the highest F1 = 96.43% and accuracy = 98.11%; the identical metrics across these thresholds indicate that no predicted probabilities lie in (0.5, 0.7], hence the same class assignments. Considering this balance and the elimination of false alarms, we adopt $\tau = 0.5$ as the primary operating point with TP=27, FP=0, FN=2, TN=77 as shown in Fig. 5. Finally, the model's AUC-ROC = 0.9978, indicating excellent threshold-independent discriminative ability.

IV. CONCLUSIONS

This study presented a supervised neural-network classifier for acoustic anomaly detection in industrial fan systems using engineered features. Exploratory and statistical analyses showed that MFCC_1 to MFCC_6, spectral centroid, spectral roll-off, and autocorrelation provide the strongest discrimination between normal and abnormal conditions ($p < 0.05$), whereas MFCC_11 to MFCC_13 and spectral contrast contribute limited additional value. A compact feed-forward network (two 64-unit ReLU layers with dropout) trained with stratified 5-fold cross-validation achieved a mean F1 of 89.9%. Threshold sweeping ($\tau \in \{0.3, 0.4, 0.5, 0.6, 0.7\}$) on the held-out test set confirmed robust behavior; the selected operating point $\tau = 0.5$ attains precision = 100%, recall = 93.10%, F1 = 96.43%, and accuracy = 98.11% (identical metrics for $\tau = 0.6-0.7$), while $\tau = 0.3$ yields higher recall at the expense of precision. The model's AUC-ROC = 0.9978 indicates near-perfect threshold-independent separability. These findings demonstrate that combining interpretable acoustic features with a lightweight neural classifier enables accurate, non-invasive anomaly detection suitable for predictive-maintenance workflows with minimal hardware overhead. This study contributes to the advancement of predictive maintenance frameworks, providing a cost-effective and non-invasive alternative to traditional sensor-based monitoring. Future work could explore the generalization of this approach to other industrial machinery and the incorporation of unsupervised learning techniques to further enhance anomaly detection in scenarios with limited labeled data.

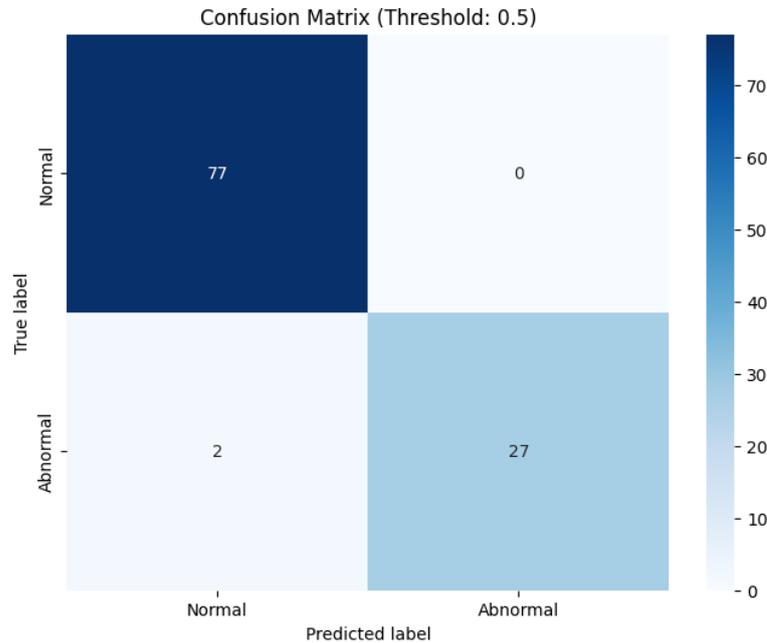


Fig. 5 Confusion matrix on the test set at the selected operating threshold $\tau=0.5$ (positive class: *Abnormal*)

REFERENCES

- [1] C.-S. A. Gong *et al.*, “Design and implementation of acoustic sensing system for online early fault detection in industrial fans,” *J. Sens.*, vol. 2018, no. 1, p. 4105208, 2018. doi: 10.1155/2018/4105208.
- [2] R. Müller, F. Ritz, S. Illium, and C. Linnhoff-Popien, “Acoustic anomaly detection for machine sounds based on image transfer learning,” *arXiv preprint arXiv:2006.03429*, 2020. doi: 10.5220/0010185800490056.
- [3] Y. Ota and M. Unoki, “Anomalous sound detection for industrial machines using acoustical features related to timbral metrics,” *IEEE Access*, vol. 11, pp. 70884–70897, 2023. doi: 10.1109/ACCESS.2023.3294334.
- [4] D. C. Sulaiman and T. M. S. Mulyana, “Web-based writing learning application of basic Hanacaraka using convolutional neural network method,” *Ultimatics: J. Tek. Inform.*, vol. 15, no. 1, pp. 28–34, 2023. doi: 10.31937/ti.v15i1.2993.
- [5] B. Hakim, F. J. Kaunang, C. Susanto, J. Salim, and R. Indradjaja, “Implementasi machine learning dalam pengelompokan musik menggunakan algoritma K-means clustering,” *Idealis: Indones. J. Inf. Syst.*, vol. 8, no. 1, pp. 74–83, 2025. doi: 10.36080/idealis.v8i1.3357.
- [6] L. Y. Teck and A. P. Thenata, “Stock price prediction using TCN-GAN hybrid model,” *Sinkron: J. Penelit. Tek. Inform.*, vol. 9, no. 1, pp. 106–114, 2025. doi: 10.33395/sinkron.v9i1.14246.
- [7] M. Fernandes, J. M. Corchado, and G. Marreiros, “Machine learning techniques applied to mechanical fault diagnosis and fault prognosis in the context of real industrial manufacturing use-cases: a systematic literature review,” *Appl. Intell.*, vol. 52, no. 12, pp. 14246–14280, 2022. doi: 10.1007/s10489-022-03344-3.
- [8] R. Marino, C. Wisultschew, A. Otero, J. M. Lanza-Gutierrez, J. Portilla, and E. de la Torre, “A machine-learning-based distributed system for fault diagnosis with scalable detection quality in industrial IoT,” *IEEE Internet Things J.*, vol. 8, no. 6, pp. 4339–4352, 2020. doi: 10.1109/IIOT.2020.3026211.
- [9] M. K. Gourisaria, R. Agrawal, M. Sahni, and P. K. Singh, “Comparative analysis of audio classification with MFCC and STFT features using machine learning techniques,” *Discover Internet Things*, vol. 4, no. 1, p. 1, 2024. doi: 10.1007/s43926-023-00049-y.
- [10] M. S. Sidhu, N. A. A. Latib, and K. K. Sidhu, “MFCC in audio signal processing for voice disorder: a review,” *Multimed. Tools Appl.*, vol. 84, no. 10, pp. 8015–8035, 2025. doi: 10.1007/s11042-024-19253-1.
- [11] S. D. D. Anton, S. Sinha, and H. D. Schotten, “Anomaly-based intrusion detection in industrial data with SVM and random forests,” in *Proc. Int. Conf. Softw., Telecommun. Comput. Netw. (SoftCOM)*, Split, Croatia: IEEE, 2019, pp. 1–6. doi: 10.23919/SOFTCOM.2019.8903672.
- [12] M. Aly and M. H. Behiry, “Enhancing anomaly detection in IoT-driven factories using logistic boosting, random forest, and SVM: A comparative machine learning approach,” *Sci. Rep.*, vol. 15, no. 1, p. 23694, 2025. doi: 10.1038/s41598-025-08436-x.
- [13] S.-M. Kim and Y. S. Kim, “Enhancing sound-based anomaly detection using deep denoising autoencoder,” *IEEE Access*, vol. 12, pp. 84323–84332, 2024. doi: 10.1109/ACCESS.2024.3414435.
- [14] L. Ruff *et al.*, “Deep semi-supervised anomaly detection,” *arXiv preprint arXiv:1906.02694*, 2019.
- [15] P. Chennai Viswanathan *et al.*, “Deep learning for enhanced fault diagnosis of monoblock centrifugal pumps: Spectrogram-based analysis,” *Machines*, vol. 11, no. 9, p. 874, 2023. doi: 10.3390/machines11090874.
- [16] N. Dutta, P. Kaliannan, and U. Subramaniam, “Application of machine learning algorithm for anomaly detection for industrial pumps,” in *Mach. Learn. Algorithms Ind. Appl.*, Singapore: Springer, 2020, pp. 237–263. doi: 10.1007/978-3-030-50641-4_14.
- [17] D. Cabrera, R. Medina, M. Cerrada, R.-V. Sanchez, E. Estupinan, and C. Li, “Improved Mel frequency cepstral coefficients for compressors and pumps fault diagnosis with deep learning models,” *Appl. Sci.*, vol. 14, no. 5, p. 1710, 2024. doi: 10.3390/app14051710.
- [18] L. Ttito Ugarte and F. Bernardini, “An overview on the use of machine learning algorithms for identifying anomalies in industrial valves,” in *World Conf. Inf. Syst. Technol.*, Cham: Springer, 2024, pp. 3–12. doi: 10.1007/978-3-031-60215-3_1.
- [19] B. R. Navada, V. Sravani, and S. K. Venkata, “Enhancing industrial valve diagnostics: Comparison of two preprocessing methods on the performance of a stiction detection method using an artificial neural network,” *Appl. Syst. Innov.*, vol. 7, no. 6, p. 104, 2024. doi: 10.3390/asi7060104.
- [20] K. J. Folz and H. M. Gomes, “An investigation of machine learning strategies for electric motor anomaly detection using vibration and audio signals,” *Eng. Comput. (Swansea)*, vol. 42, no. 2, pp. 465–487, 2025. doi: 10.1108/EC-03-2024-0206.
- [21] E. Yun and M. Jeong, “Acoustic feature extraction and classification techniques for anomaly sound detection in the electronic motor of automotive EPS,” *IEEE Access*, 2024. doi: 10.1109/ACCESS.2024.3471169.
- [22] H. Purohit *et al.*, “MIMII dataset: Sound dataset for malfunctioning industrial machine investigation and inspection,” *arXiv preprint arXiv:1909.09347*, 2019. doi: 10.33682/m76f-d618.

Angelina Pramana Thenata. Akademisi di bidang Informatika dan aktif meneliti topik *machine learning*, *software engineering*, dan *system analysis*. Karya ilmiahnya mendukung pengambilan keputusan berbasis data dan sistem cerdas.

Ranny. Menyelesaikan S3 Teknik Elektro dan Informatika di Institut Teknologi Bandung pada 2024 dengan latar belakang pendidikan S1 dan S2 di bidang Teknik Informatika/Ilmu Komputer. Publikasinya meliputi *data science*, *artificial intelligence*, rekayasa perangkat lunak, dan pemanfaatan teknologi untuk pemberdayaan masyarakat.

Bhustomy Hakim. Menempuh pendidikan Master di Harbin Institute of Technology jurusan Computer Science. Saat ini aktif mengajar dan melakukan penelitian di bidang ilmu *data science* dan pengembangan sistem.

Fergie Joanda Kaunang. Menyelesaikan studi S1 di bidang Informatika serta melanjutkan S2 dengan fokus *bioinformatics* dan *machine learning*. Saat ini aktif sebagai akademisi dengan minat riset pada *data mining*, *machine learning*, *bioinformatics*, dan *system analysis*.