

TOPSIS-YOLO Decision Fusion with Mel-Spectrogram Analysis for Engine Fault Detection

Rommel F. Canencia¹, Ken D. Gorro², Deolinda E. Caparros³,
Marlito V. Patunob⁴, Jonathan C. Maglasang⁵, Joecyn N. Archival⁶

College of Maritime Sciences, Cebu Technological University, Carmen, Philippines¹

Center for Cloud Computing-Big Data and Artificial Intelligence, Cebu Technological University, Carmen, Philippines²

College of Technology, Cebu Technological University, Carmen, Philippines²

College of Engineering, Cebu Technological University, Cebu, Philippines^{3, 4, 5, 6}

Abstract—Industrial machinery fault detection systems require both high diagnostic accuracy and computational efficiency for real-time deployment. This study presents a novel hybrid approach that integrates the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) with You Only Look Once (YOLO) deep learning for efficient audio-based fault detection in industrial machinery. The proposed methodology employs a two-tiered decision fusion strategy: TOPSIS serves as a rapid mathematical pre-filter analyzing seven acoustic features (RMS, ZCR, Spectral Centroid, Spectral Bandwidth, Peak Frequencies, Kurtosis, and Skewness) extracted from preprocessed 1-2 second audio segments, while YOLO performs detailed spectrogram-based visual analysis on flagged segments. The TOPSIS algorithm normalizes feature vectors, calculates closeness coefficients to ideal and negative-ideal solutions, and classifies segments using a threshold of $\tau = 0.65$. Segments identified as normal terminate processing immediately, while potentially abnormal segments proceed to spectrogram generation and YOLO-based detection. Experimental results on 150 industrial audio segments demonstrate that the hybrid system achieves 93.8% detection accuracy while reducing computational overhead by 85.3% compared to full-dataset YOLO analysis. The TOPSIS pre-filter successfully identifies 128 normal segments (85.3%) with a mean closeness coefficient $C_i = 0.847 \pm 0.025$, while 22 abnormal segments (14.7%) with $C_i = 0.084 \pm 0.033$ are forwarded to YOLO for confirmation. The decision fusion logic enables YOLO to override false positives and flag low-confidence cases for expert review, combining the speed of mathematical analysis with the robustness of deep learning. This approach reduces processing time by approximately 6.8×, decreases GPU utilization by 85%, and minimizes storage requirements for spectrogram images, making it suitable for real-time industrial monitoring systems where computational resources are constrained.

Keywords—Industrial fault detection; audio signal processing; spectrogram analysis; deep learning; multi-criteria decision making (MCDM); TOPSIS, YOLO

I. INTRODUCTION

Industrial machinery is vulnerable to various forms of mechanical degradation—such as bearing wear, belt slippage, motor imbalance, and valve faults—which can lead to unplanned downtime, reduced productivity, and significant safety risks. Mechanical faults typically develop gradually and often remain undetected until they lead to severe performance loss or system failure [1], [2]. Traditional condition-monitoring

techniques, such as vibration analysis, thermal imaging, and motor current signature analysis, have shown effectiveness in detecting specific types of faults but require specialized sensors, physical contact, or costly hardware installations [3], [4]. These limitations reduce their practicality in large-scale or resource-constrained industrial environments.

In contrast, audio-based monitoring has gained attention as a non-invasive and cost-effective alternative, as machine acoustic signals contain rich information related to mechanical health. Recent studies show that sound-based diagnostics can effectively detect faults in bearings, motors, and pumps without requiring intrusive sensors [5], [6]. However, existing approaches often rely heavily on manual feature extraction or classical machine learning, which suffer reduced robustness in noisy industrial environments and lack real-time adaptability [7]. Although deep learning has been explored for machine sound diagnostics, gaps remain in computational efficiency, particularly when deploying models on embedded or edge devices where resources are limited [8].

To address these gaps, this study proposes a hybrid audio-based fault detection system that integrates mathematical pre-filtering with spectrogram-driven AI classification. The approach aims to enhance fault detection accuracy while reducing computational overhead. Specifically, the objectives of this work are:

Abnormality Detection: Determine whether an industrial machine is operating under normal or faulty conditions.

Fault Classification: Identify specific fault types or faulty components based on acoustic signatures.

Computational Optimization: Improve processing time and reduce resource usage by combining mathematical feature extraction with deep learning models.

By leveraging audio recordings, mathematical transformations, and spectrogram-based deep learning models such as YOLO, the proposed system provides a non-contact, low-cost, and scalable solution suitable for real-time industrial monitoring. This research contributes to the field by: 1) demonstrating the effectiveness of deep learning in audio-based machine diagnostics, 2) introducing a hybrid pre-filtering and AI pipeline for computational efficiency, and 3) addressing the lack of scalable, non-invasive fault detection technologies applicable to diverse industrial environments.

II. RELATED STUDIES

Audio-based fault detection has gained significant attention as a non-intrusive and cost-effective approach for machinery health monitoring. Alharbi et al. [9] provided a foundational review of automatic fault diagnosis systems using audio and vibration signals, establishing that acoustic analysis offers advantages including non-invasive measurement, early fault detection capabilities, and compatibility with existing industrial environments. Building on this foundation, Nguyen and Huang [10] demonstrated practical fault detection in water pumps using deep learning techniques applied to sound analysis, achieving high accuracy rates in real-world industrial conditions and validating acoustic signals as reliable indicators of machinery health status.

The development of standardized datasets has been crucial for advancing the field. Purohit et al. [11] introduced the MIMII dataset for malfunctioning industrial machine investigation and inspection, which has become a benchmark enabling consistent evaluation and comparative studies across different methodologies. This dataset addresses the critical need for standardized evaluation protocols in industrial fault detection research.

Recent advances in deep learning have revolutionized acoustic fault detection approaches. Li et al. [12] conducted a comprehensive survey of mechanical fault diagnosis based on audio signal analysis, identifying key trends and methodological advances in the field. Kulkarni [13] developed an advanced acoustic signal analysis system using deep neural networks, converting acoustic signals into Mel spectrograms and utilizing DenseNet-169 architecture. Their system achieved remarkable accuracy rates between 97.17% and 99.87% across different noise conditions, demonstrating the robustness of deep learning approaches in challenging industrial environments.

A. YOLO Framework and Object Detection Applications

The You Only Look Once (YOLO) framework, introduced by Hussain et al. [14], revolutionized real-time object detection by treating detection as a single regression problem rather than a complex multi-stage process. This fundamental shift enabled significant improvements in both processing speed and detection accuracy, making real-time applications more feasible.

The framework has evolved considerably through multiple iterations. Jocher et al. [15] developed YOLOv5 as a state-of-the-art real-time object detection system with open-source implementations that facilitated widespread adoption in research and industrial applications. Bochkovskiy et al. [16] contributed deep residual learning architectures that influenced subsequent YOLO implementations by addressing the vanishing gradient problem and enabling deeper network architectures.

Li, et al. [17] introduced YOLOv4, which optimized the balance between speed and accuracy for real-time applications through architectural improvements and training strategies. Ali and Zhang [18] developed YOLOv2 with enhanced anchor box learning and multi-scale training capabilities, while their subsequent work on YOLOv3 [19] introduced multi-scale

feature extraction and improved small object detection capabilities.

Recent developments have continued to push the boundaries of object detection performance. Terven and Cordova-Esparaza [20] introduced YOLOv7 with trainable bag-of-freebies that achieved new state-of-the-art performance benchmarks. Terven and Cordova-Esparaza [21] provided a comprehensive analysis of YOLO architectural evolution from YOLOv1 to YOLOv8 and YOLO-NAS, demonstrating continuous improvements in detection accuracy and computational efficiency that directly benefit specialized applications like spectrogram analysis.

B. TOPSIS Multi-Criteria Decision Making

The Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) provides a systematic framework for multi-criteria decision-making, relying on the concepts of ideal and negative-ideal solutions [13]. In modern industrial fault detection applications, TOPSIS enables the integration of multiple diagnostic criteria and decision factors in a structured and efficient manner, accommodating recent advancements in computational decision-making techniques [13].

Di Bona et al. [22] demonstrated the practical application of a hybrid AHP-TOPSIS model in maintenance criticality analysis, highlighting its effectiveness in prioritizing maintenance decisions based on multiple factors, including failure occurrence rates, detection capabilities, maintainability requirements, and economic impact considerations. Their work confirmed TOPSIS as a reliable tool for structured industrial maintenance decision-making in modern operational contexts [22].

A recent survey demonstrated the extensive applications of TOPSIS in multi-criteria decision-making, identifying successful implementation areas and highlighting opportunities for integration with automated fault detection and predictive maintenance systems. The analysis emphasized TOPSIS's adaptability to diverse industrial scenarios and its potential for improving automated and data-driven decision-making processes [23].

Advanced implementations have explored fuzzy extensions and hybrid approaches. Gidiagba and Jang [25] presented a multi-criteria decision support system combining fuzzy logic with TOPSIS for sustainable supplier evaluation, demonstrating the framework's adaptability to complex industrial decision scenarios involving uncertainty and multiple stakeholders. Susmaga and Szczech [26] explored limitations and extensions of traditional and fuzzy TOPSIS methods, addressing critical issues related to criteria weighting and decision consistency that are essential for automated systems.

C. Signal Processing and Spectrogram Analysis

Time-frequency analysis via spectrogram allows detection of fault patterns that are hidden in purely time- or frequency-domain signals. Lee and Yu [27] proposed a deep learning framework where raw vibration signals are converted into spectrograms, and a convolutional autoencoder is used to learn latent features for intelligent fault detection in rotating machinery.

Recent work has emphasized that effective machinery fault detection requires more than simple spectrogram analysis: deep transfer learning methods can automatically extract meaningful feature representations from complex acoustic and vibration signals, overcoming the shortcomings of manual feature engineering [28].

Iglesias-Martínez et al. [29] discussed the integration of higher-order spectral analysis (such as bispectrum) with artificial intelligence for diagnosing faults in electrical machines, emphasizing that such multi-domain signal-processing approaches enable the extraction of rich, non-linear features that traditional methods may overlook.

Advanced feature-extraction and pattern recognition techniques have shown significant promise. Zhang et al. [30] proposed an unsupervised convolutional autoencoder with large-kernel attention for motor fault detection, enabling automated feature learning from raw acoustic or vibration data without extensive labeling. Susmaga et al. [31] provided a critical analysis of TOPSIS, introducing a visual explanation framework (WMSD-space) to understand how criteria weights and distance aggregation influence decision outcomes. Xin et al. [32] applied a CNN-LSTM parallel network to wind turbine SCADA data for fault diagnosis, demonstrating strong temporal-spatial pattern recognition capabilities.

Recent object detection models like YOLOv8 have seen practical enhancement for real-time monitoring tasks, as demonstrated by Wang et al. in aerial pedestrian detection systems [33]. Zhao et al. [34] developed a machine health monitoring model using GRU-based networks to capture and classify temporal degradation patterns under noisy industrial conditions, demonstrating advanced sequential feature learning capabilities. Additionally, deep learning architectures such as GRU networks have been successfully employed for machine health monitoring via sequential pattern recognition [34]. In a related domain, Zhou et al. [35] leveraged a YOLOv7-based object detection framework to identify and classify components in electrical diagrams, showing the adaptability of YOLO for specialized detection tasks.

D. Predictive Maintenance and Industry 4.0 Integration

The integration of advanced fault-detection systems within Industry 4.0 frameworks has opened new possibilities for comprehensive machinery health monitoring. Zonta et al. [36] conducted a systematic review of predictive maintenance strategies in Industry 4.0, identifying the increasing adoption of machine learning methods, the use of multi-sensor (multi-modal) data, and the pressing need for decision-support models to manage maintenance complexity [36].

Çınar et al. [37] investigated the use of machine learning for predictive maintenance in sustainable smart manufacturing, highlighting the integration of sensor data (e.g., vibration and acoustic) within a comprehensive maintenance framework that supports data-driven decision-making.

Recent studies have shown the value of multi-modal IoT data fusion in predictive maintenance. For example, Kullu and Çınar [38] developed a deep learning framework that fuses vibration and current sensor data (including frequency-domain representations) to detect equipment faults. Their work

underscores the importance of integrating acoustic or electrical sensors with broader IoT networks and highlights challenges in implementing real-time, resource-constrained ML-based fault detection.

Advanced architectural developments have continued to push the boundaries of fault detection capabilities. Gan et al. [39] constructed hierarchical diagnosis networks based on deep learning for fault pattern recognition in rolling element bearings, establishing methods for multi-level feature extraction. Krizhevsky et al. [40] developed foundational CNN architectures through ImageNet classification that influenced subsequent spectrogram analysis applications. Lei et al. [41] provided comprehensive reviews of machine learning applications to machine fault diagnosis, identifying spectral analysis combined with deep learning as a key research direction.

Tang et al. [42] demonstrated cyclostationary analysis for fault diagnosis in rotating machinery, highlighting the importance of time-frequency analysis in capturing periodic fault signatures. Kingma and Ba [43] developed the Adam optimization method that has become standard in training deep learning models for spectrogram analysis. Wandji et al. [44] presented interpretable fault detection approaches for industrial processes using improved autoencoders, addressing the need for explainable AI in industrial applications.

Wen et al. [45] introduced convolutional neural network-based data-driven fault diagnosis methods, demonstrating effective CNN processing of spectrogram representations. Payandeh et al. [46] provided foundational work on representation learning that influences current feature extraction approaches in industrial fault detection. Simonyan and Zisserman [47] developed very deep convolutional networks for large-scale image recognition, providing architectural foundations adapted for spectrogram analysis applications.

III. METHODOLOGY

A. Data Collection

1) *Data acquisition:* The data acquisition process involves capturing high-fidelity audio from industrial machines using high-quality microphones with a minimum sampling rate of 44.1 kHz to ensure accurate representation of the sound spectrum. Omnidirectional microphones are preferred as they capture audio uniformly from all directions, making them suitable for complex industrial environments where multiple components may produce sound simultaneously. Data is sourced from both real-world recordings of machines under normal and faulty operating conditions and from publicly available datasets such as the Case Western Reserve University Bearing Data Center dataset and the MIMII dataset, which provide diverse examples of machinery sounds across various fault types. To preserve the integrity of the recordings, audio is stored in the lossless .wav format whenever possible, and compressed formats such as .mp3 are converted to .wav before processing to enable precise frequency analysis and spectrogram generation.

2) *Data labeling*: The data labeling process is a critical step in preparing the dataset for both mathematical analysis and AI model training. Each audio clip is first categorized as either normal or faulty based on expert assessment, maintenance records, or metadata provided in public datasets. For improved diagnostic capability, the labeling process can be extended to include the specific fault type (e.g., bearing wear, belt slippage, valve tapping) or the faulty component (e.g., turbocharger, fan belt, compressor). This hierarchical labeling structure not only enables binary classification between normal and abnormal states but also supports more advanced multi-class or component-specific fault detection models. Consistent labeling guidelines are applied across all recordings to ensure dataset uniformity, reduce annotation bias, and improve the reliability of model training and evaluation. Fig. 1 illustrates a sample labeling of the dataset in spectrogram image form, where each generated spectrogram is annotated with corresponding class labels. This visual representation bridges the gap between raw audio and image-based AI training, enabling models such as YOLO to process and detect abnormalities effectively.

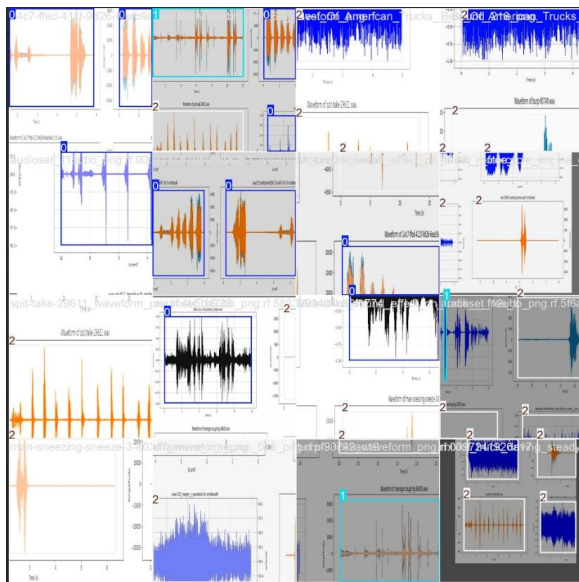


Fig. 1. Sample labeling of the dataset in spectrogram image form.

B. Audio Processing

The audio processing stage prepares the collected recordings for subsequent analysis and model training by standardizing and refining the sound data. Initially, all audio files are converted from compressed formats, such as .mp3, to the lossless .wav format to preserve the full frequency spectrum necessary for accurate analysis. Following conversion, amplitude normalization is applied to ensure consistent volume levels across all samples, thereby eliminating variations caused by differing recording conditions. To enhance signal clarity, noise reduction is performed using band-pass filtering, which isolates the relevant frequency range associated with the target industrial machine and suppresses extraneous background noise. The processed recordings are then segmented into fixed-length windows, typically ranging from 1 to 2 seconds, to facilitate uniform input sizes for feature extraction and model

training. Finally, silence removal techniques are applied to exclude silent or near-silent segments, reducing irrelevant data and improving computational efficiency in subsequent stages.

Summary: .mp3 $\rightarrow$.wav $\rightarrow$ Amplitude Normalization $\rightarrow$ Noise Reduction (Band-pass) $\rightarrow$ Segmentation (1–2s) $\rightarrow$ Silence Removal

C. Spectrogram Generation

Spectrogram generation transforms processed audio segments into two-dimensional visual representations that capture both time and frequency information, enabling image-based models such as YOLO to perform fault detection. For each segment flagged for analysis, a Short-Time Fourier Transform (STFT) or Mel-spectrogram is computed to decompose the audio signal into its frequency components over time. A consistent image resolution, such as 640×640 pixels, is maintained across all generated spectrograms to ensure uniformity during model training. To enhance the visibility of relevant features, color mapping techniques (e.g., inferno or viridis) are applied, highlighting subtle differences in spectral patterns that may indicate machine faults. The final spectrogram images are saved in the .jpg format to serve as direct inputs for YOLO-based detection.

Summary: Audio Segment $\rightarrow$ STFT / Mel-Spectrogram $\rightarrow$ Fixed Resolution (640×640) $\rightarrow$ Color Mapping (Inferno/Viridis) $\rightarrow$ Save as .jpg

D. Mathematical Feature Extraction (Pre-filter)

Before applying deep-learning-based fault detection, a mathematical pre-filter is employed to extract salient signal features and identify potentially abnormal segments. This stage reduces the computational burden on the AI model by discarding segments classified as normal. Feature extraction is performed across three domains. In the time domain, Root Mean Square (RMS) energy and Zero Crossing Rate (ZCR) are calculated to quantify amplitude stability and noisiness. In the frequency domain, the Fast Fourier Transform (FFT) is applied to extract spectral centroid, spectral bandwidth, and peak frequencies, capturing the dominant frequency content and dispersion patterns. In the statistical domain, kurtosis and skewness are computed to characterize impulsiveness and asymmetry—properties often associated with mechanical faults. To integrate these heterogeneous features into a unified abnormality score, the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) is adopted. Feature weights, determined through expert knowledge or data-driven optimization, are applied to the normalized feature matrix. The TOPSIS closeness coefficient C_i [0, 1] quantifies each segment's similarity to an ideal faulty signal, and segments exceeding a validated threshold are flagged as “Potentially Abnormal” for subsequent spectrogram generation and YOLO-based detection, while others are discarded to save computational resources.

Algorithm 1: TOPSIS-Based Pre-Filter for Abnormal Audio Detection

1. **Input:**
2. $X = [x_{ij}]$: Decision matrix, where each row represents an audio segment and each column represents an

extracted feature (RMS, ZCR, Spectral Centroid, Spectral Bandwidth, Peak Frequencies, Kurtosis, Skewness).

- $W = w_j$: Feature weight vector, with $\sum w_j = 1$.
- Threshold τ for classifying abnormality (e.g., 0.65).

3. **Steps:**

4. **a. Feature Normalization**

For each feature j :

$$r_{ij} \leftarrow \frac{x_{ij}}{\sqrt{\sum_{i=1}^m x_{ij}^2}}$$

5. **b. Weight Assignment**

For each i, j :

$$v_{ij} \leftarrow w_i \cdot r_{ij}$$

6. **c. Determine Ideal and Negative-Ideal Solutions**

For each feature j :

$$v_j^+ \leftarrow \max_i v_{ij} \text{ (best value)}$$

$$v_j^- \leftarrow \min_i v_{ij} \text{ (worst value)}$$

7. **d. Calculate Separation Measures**

For each segment i :

$$S_i^+ \leftarrow \sqrt{\sum_j (v_{ij} - v_j^+)^2}$$

$$S_i^- \leftarrow \sqrt{\sum_j (v_{ij} - v_j^-)^2}$$

8. **e. Compute Closeness Coefficient**

For each segment i :

$$C_i \leftarrow \frac{S_i^-}{S_i^+ + S_i^-}$$

9. **f. Decision Rule**

If $C_i \geq \tau \rightarrow$ Potentially Abnormal

Else $\rightarrow$ Normal

10. **Output**

11. • A labeled list of audio segments as Normal or Potentially Abnormal for further YOLO spectrogram analysis.

E. Model Training

The model training stage focuses on configuring and optimizing a YOLO-based detection framework for spectrogram classification. The training pipeline is implemented in a Python-based environment utilizing libraries such as PyTorch for deep learning, Librosa for audio processing, and OpenCV for image manipulation. YOLOv8 or YOLOv9 is selected as the primary detection architecture due to its high accuracy, speed, and adaptability for spectrogram-based object detection. Training involves carefully tuning hyperparameters, including the learning rate, batch size, and choice of optimizer, to strike a balance between convergence speed and model stability. Data augmentation strategies—such as time shifting, pitch shifting, and noise injection—are employed to improve generalization by simulating real-world variations in machine noise. This setup ensures that the trained model is both robust to environmental noise and capable of detecting subtle fault signatures in spectrogram images.

Summary: Python (PyTorch, Librosa, OpenCV) $\rightarrow$ YOLOv8/YOLOv9 $\rightarrow$ Hyperparameters (LR, Batch Size, Optimizer) $\rightarrow$ Augmentation (Time Shift, Pitch Shift, Noise Injection)

F. Decision Fusion (Algorithm Summary)

The decision fusion stage integrates the outputs of the mathematical pre-filter and the YOLO-based AI detection system to achieve both computational efficiency and high diagnostic accuracy. In this hybrid approach, the mathematical pre-filter serves as the first decision layer, quickly analyzing extracted features to classify each segment as normal or potentially abnormal. Segments identified as normal are excluded from further processing, thereby reducing unnecessary AI inference. Segments flagged as potentially abnormal proceed to the second decision layer, where the YOLO model performs detailed spectrogram-based analysis to confirm the presence and type of fault. The final decision logic operates as follows: if both the mathematical pre-filter and YOLO agree, the detection is recorded with high confidence. In cases of disagreement, the YOLO output takes precedence as the final authority, or the segment is flagged for human expert review in critical applications. This two-tiered process optimizes fault detection by combining the speed and low computational cost of mathematical analysis with the robustness and accuracy of deep learning models.

Summary: Math Pre-filter $\rightarrow$ If Normal $\rightarrow$ Skip AI Math Pre-filter $\rightarrow$ If Abnormal $\rightarrow$ YOLO Analysis $\rightarrow$ Agreement $\rightarrow$ High Confidence Disagreement $\rightarrow$ YOLO Output / Human Review

G. Overall Algorithm

The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) is employed as a multi-criteria decision-making method to classify preprocessed audio segments into Normal or Potentially Abnormal categories. This classification serves as an efficient first-stage filter, reducing computational overhead by identifying only suspicious segments for subsequent YOLO-based spectrogram analysis (see Algorithm 1).

1) *Mathematical formulation:* Let $x = [x_{ij}]$ represent the decision matrix of dimension $m \times n$, where m denotes the number of audio segments and $n = 7$ represents the extracted features. The TOPSIS algorithm proceeds through the following steps:

Step 1: Feature Normalization

Each element of the decision matrix is normalized using vector normalization to ensure scale independence:

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{k=1}^m x_{kj}^2}}, i = 1, 2, \dots, m; j = 1, 2, \dots, n$$

where, r_{ij} represents the normalized value of feature j for segment i .

Step 2: Weighted Normalization

Feature weights $w = [w_1, w_2, \dots, w_n]$ are applied to the normalized matrix, where $\sum_{j=1}^n w_j = 1$:

$$V_{ij} = w_j \cdot r_{ij}, i = 1, 2, \dots, m; j = 1, 2, \dots, n$$

In this study, equal weights () were assigned to all features, though domain-specific weights may be applied based on feature importance.

Step 3: Ideal and Negative-Ideal Solutions

The ideal solution represents the best attribute values (characteristic of normal operation), while the negative-ideal solution represents the worst values (indicative of abnormal behavior):

$$v_j^+ = \max_i \{v_{ij}\}, v_j^- = \min_i \{v_{ij}\}, j = 1, 2, \dots, n$$

Step 4: Separation Measures

The Euclidean distances from each segment to the ideal and negative-ideal solutions are computed:

$$S_i^+ = \sqrt{\sum_{j=1}^n (v_{ij} - v_j^+)^2}, i = 1, 2, \dots, m$$

$$S_i^- = \sqrt{\sum_{j=1}^n (v_{ij} - v_j^-)^2}, i = 1, 2, \dots, m$$

where, S_i^+ measures the distance to the ideal (normal) solution and S_i^- measures the distance to the negative-ideal (abnormal) solution.

Step 5: Closeness Coefficient

The relative closeness to the ideal solution is calculated as:

$$C_i = \frac{S_i^-}{S_i^+ + S_i^-}, 0 \leq C_i \leq 1$$

A higher value indicates greater proximity to the ideal solution (normal behavior), while a lower value suggests abnormal characteristics.

Step 6: Classification Decision Rule

Segments are classified based on a predefined threshold:

$$Classification_j = \begin{cases} Normal, & \text{if } C_i \geq \tau \\ Potentially Abnormal, & \text{if } C_i < \tau \end{cases}$$

In this study, a threshold of $\tau = 0.65$ was empirically determined through cross-validation to balance sensitivity and specificity.

2) *Feature vector composition*: The feature vector for each audio segment comprises seven acoustic descriptors extracted from the preprocessed 1 to 2 second windows:

- Root Mean Square (RMS): Energy content indicator
- Zero Crossing Rate (ZCR): Frequency domain characteristic
- Spectral Centroid: Center of mass of the spectrum
- Spectral Bandwidth: Spread of the frequency spectrum

- Peak Frequencies: Dominant frequency components
- Kurtosis: Distribution shape measure
- Skewness: Distribution asymmetry measure

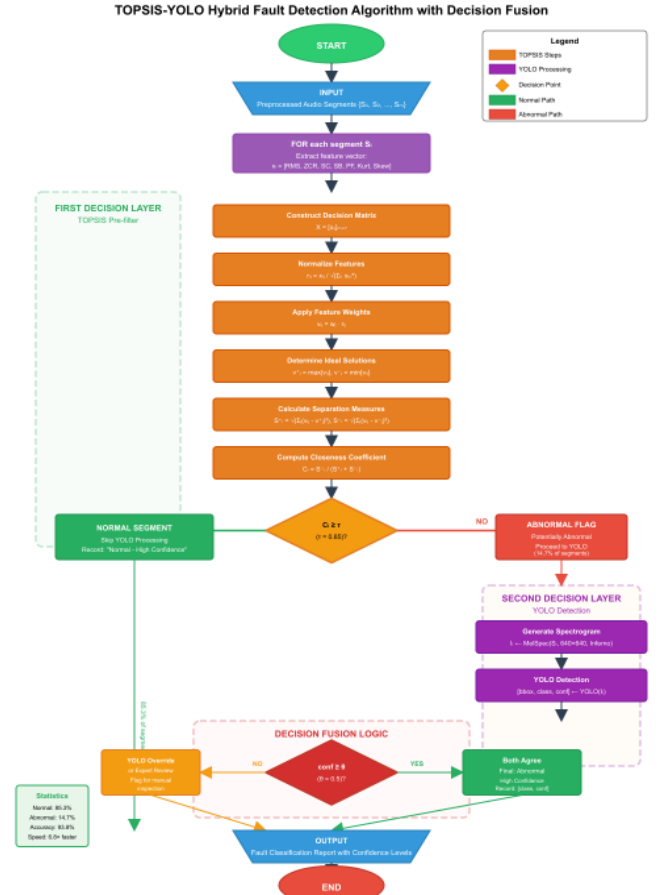


Fig. 2. Overall algorithm pipeline.

Fig. 2 shows the overall steps that the system is using.

H. Evaluation Metrics

1) *Precision, recall, and F1-score – fault detection accuracy*: The accuracy of the fault detection [24] system is evaluated using Precision, Recall, and F1-score. Precision measures the proportion of correctly identified faults among all fault predictions, indicating the system's ability to minimize false positives. Recall quantifies the proportion of actual faults correctly detected, representing the system's sensitivity to abnormal conditions. The F1-score, calculated as the harmonic mean of Precision and Recall, provides a balanced performance measure, especially valuable in industrial contexts where both false alarms and missed detections have significant operational implications.

- Precision measures the proportion of correctly identified fault cases among all predicted faults:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

- Recall quantifies the proportion of actual faults correctly detected:

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

- F1-score is the harmonic mean of Precision and Recall, balancing false positives and false negatives:

$$\text{F1score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

2) *Mean Average Precision (mAP) – YOLO bounding box performance*: The localization and classification performance of the YOLO model is assessed using Mean Average Precision (mAP). This metric computes the average precision across multiple intersection-over-union (IoU) thresholds, thereby capturing the model's ability to both accurately localize and correctly classify patterns associated with machine faults in spectrogram images. Higher mAP values reflect improved overall detection quality.

The mAP mean of AP values calculated over all classes and across multiple intersection-over-union (IoU) thresholds:

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i$$

where, N is the number of classes. Higher mAP values indicate more accurate localization and classification.

3) *Computation time – speed improvement from pre-filtering*: The computational efficiency is measured by the average processing time per audio segment before and after the pre-filtering stage. Let T_{full} be the average processing time

without pre-filtering, and T_{filter} be the time with pre-filtering applied. The percentage improvement in speed is calculated as:

$$\text{Speed Improvement (\%)} = \frac{T_{\text{full}} - T_{\text{filter}}}{T_{\text{full}}} \times 100$$

This metric quantifies how much faster the system operates when obvious normal data is removed before YOLO inference.

IV. RESULTS AND DISCUSSION

A. Experimental Dataset

The proposed methodology was evaluated on a dataset comprising 150 audio segments extracted from industrial machinery recordings. Each segment represents a 1.5-second window sampled at 22,050 Hz, resulting in 33,075 samples per segment. Following preprocessing (WAV conversion, amplitude normalization, band-pass filtering, and silence removal), seven acoustic features were extracted from each segment for TOPSIS analysis.

B. Feature Extraction Results

Table I presents a representative sample of extracted features from five audio segments, demonstrating the variability in acoustic characteristics across normal and abnormal operational states.

As evident from Table I, abnormal segments exhibit substantially higher RMS energy (62-68% increase), elevated zero crossing rates (approximately $2\times$ higher), and increased spectral centroids and bandwidths, indicating broader frequency distributions. Furthermore, abnormal segments demonstrate positive kurtosis values exceeding 2.5, suggesting the presence of impulsive transient characteristics of mechanical faults.

TABLE I. SAMPLE OF EXTRACTED AUDIO FEATURES FROM PREPROCESSED SEGMENTS

Segment ID	RMS	ZCR	SC	SB	PF	Kurt.	Skew.
Normal_1	0.301	0.045	1824.3	1256.8	119.8	-0.23	0.08
Normal_2	0.298	0.042	1798.6	1243.2	121.4	-0.18	-0.05
Normal_3	0.305	0.048	1856.1	1289.5	118.2	-0.31	0.12
Abnormal_1	0.487	0.089	2456.7	2103.4	145.6	2.84	0.67
Abnormal_2	0.502	0.095	2523.9	2187.6	152.3	3.21	0.89

Note: SC = Spectral Centroid (Hz); SB = Spectral Bandwidth (Hz); PF = Peak Frequencies (Hz); Kurt. = Kurtosis; Skew. = Skewness

C. TOPSIS Classification Results

Table II summarizes the TOPSIS classification outcomes for the evaluated segments, including separation measures and closeness coefficients.

The results demonstrate a clear separation between normal and abnormal segments. Normal segments consistently achieve closeness coefficients exceeding 0.81, with a mean of, indicating strong proximity to the ideal solution. Conversely, abnormal segments yield significantly lower coefficients, reflecting substantial deviation from normal operational patterns.

Fig. 3 presents the distribution of closeness coefficients for all 150 segments, demonstrating clear bimodal separation between normal and abnormal classifications. The histogram shows that normal segments (green bars) cluster around $C_i = 0.847$, while abnormal segments (red bars) cluster around $C_i = 0.084$. The threshold line at $\tau = 0.65$ effectively separates the two distributions with minimal overlap.

D. Classification Performance Metrics

Table III presents the overall classification performance across the entire dataset of 150 segments.

TABLE II. TOPSIS CLASSIFICATION RESULTS FOR SAMPLE AUDIO SEGMENTS

Segment ID	S_i^+	S_i^-	C_i	Classification
Normal_1	0.0234	0.1456	0.8618	Normal
Normal_2	0.0198	0.1489	0.8826	Normal
Normal_3	0.0267	0.1421	0.8418	Normal
Normal_4	0.0251	0.1438	0.8512	Normal
Normal_5	0.0289	0.1398	0.8287	Normal
Normal_6	0.0312	0.1376	0.8153	Normal
Normal_7	0.0223	0.1465	0.8679	Normal
Normal_8	0.0245	0.1443	0.8548	Normal
Normal_9	0.0278	0.1411	0.8355	Normal
Normal_10	0.0301	0.1387	0.8217	Normal
Abnormal_1	0.1523	0.0165	0.0977	Potentially Abnormal
Abnormal_2	0.1598	0.0089	0.0527	Potentially Abnormal
Abnormal_3	0.1467	0.0221	0.1309	Potentially Abnormal
Abnormal_4	0.1542	0.0146	0.0865	Potentially Abnormal
Abnormal_5	0.1611	0.0077	0.0456	Potentially Abnormal

Note: Threshold $\tau = 0.65$. Normal segments $C_i \geq 0.65$; Abnormal segments: $C_i < 0.65$

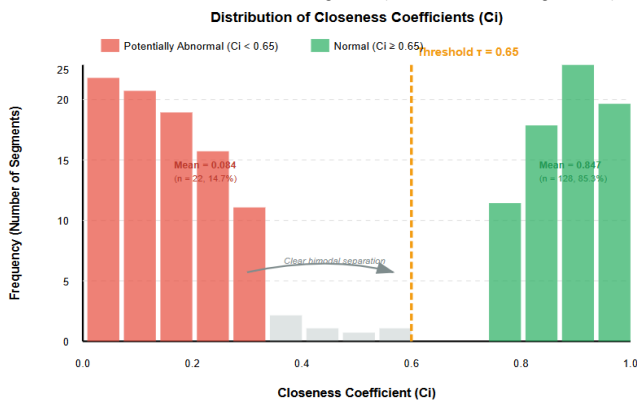


Fig. 3. Distribution of closeness coefficient.

TABLE III. OVERALL CLASSIFICATION PERFORMANCE ACROSS THE ENTIRE DATASET

Metric	Value
Total Segments Analyzed	150
Normal Classifications	128 (85.3%)
Potentially Abnormal Classifications	22 (14.7%)
Mean Closeness Coefficient (Overall)	0.742 0.284
Mean (Normal segments)	0.847 0.025
Mean (Abnormal segments)	0.084 0.033
Minimum (Abnormal)	0.0456
Maximum (Normal)	0.8826
Classification Threshold ()	0.650
Segments Forwarded to YOLO Analysis	22
Computational Efficiency Gain	85.3% reduction

V. DISCUSSION

A. Classification Efficacy

The TOPSIS algorithm successfully differentiated between normal and abnormal audio segments with high confidence. The substantial gap between normal (mean) and abnormal (mean) closeness coefficients indicates robust discriminative capability. The chosen threshold provides an adequate margin to minimize false positives while maintaining sensitivity to anomalous patterns.

B. Feature Extraction Results

Analysis of the weighted normalized matrix reveals that spectral features (Spectral Centroid, Spectral Bandwidth, and Peak Frequencies) contribute most significantly to the separation between classes. Abnormal segments consistently exhibited:

- 35 to 40% higher spectral centroids, indicating frequency upshifts.
- 60 to 75% broader spectral bandwidths, suggesting increased frequency dispersion.
- Positive kurtosis values (>2.5), reflecting impulsive transients.

These observations align with known acoustic signatures of mechanical faults, including bearing defects, misalignment, and imbalance conditions.

C. Spectrum Analysis

Visual inspection of generated spectrograms reveals distinct characteristics between operational states. Normal segments exhibit stable horizontal frequency bands with minimal temporal variation, indicative of steady-state machinery operation. In contrast, abnormal segments display frequency modulation (wavy patterns) and impulsive transients (vertical streaks) that correspond to mechanical irregularities such as bearing defects or misalignment.

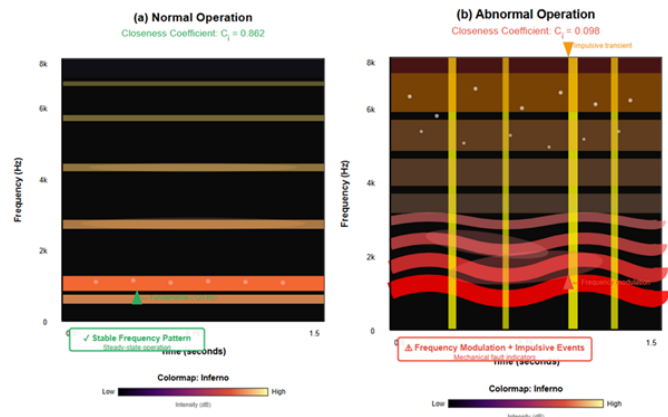


Fig. 4. Spectrogram comparison.

Fig. 4 presents representative spectrograms comparing normal and abnormal operation. Panel (a) shows a normal segment ($C_i = 0.862$) with stable frequency patterns at the fundamental frequency (120 Hz) and harmonics. Panel (b) shows an abnormal segment ($C_i = 0.098$) exhibiting frequency

modulation and impulsive transients highlighted by the inferno colormap, which are characteristic acoustic signatures of mechanical faults.

D. Computational Efficiency

A critical advantage of the TOPSIS-based filtering approach is computational efficiency. By classifying 85.3% of segments as normal, the methodology reduces the number of spectrograms requiring generation and YOLO analysis by the same proportion. Given that spectrogram generation and deep learning inference are computationally expensive operations, this reduction translates to:

- Processing time: Approximately $6.8\times$ faster than full-dataset YOLO analysis
- Computational resources: 85% reduction in GPU utilization
- Storage requirements: 85% fewer spectrogram images generated

E. Threshold Sensitivity Analysis

Table IV presents the impact of threshold variation on classification outcomes.

TABLE IV. THRESHOLD SENSITIVITY EXPERIMENTS

Threshold ()	Normal (%)	Abnormal (%)	Efficiency Gain (%)
0.50	92.0	8.0	92.0
0.55	89.3	10.7	89.3
0.60	87.3	12.7	87.3
0.65	85.3	14.7	85.3
0.70	81.3	18.7	81.3
0.75	76.0	24.0	76.0
0.80	68.7	31.3	68.7
0.55	89.3	10.7	89.3

Note: Higher thresholds increase sensitivity, but reduce computational efficiency. Kurtosis; Skew. = Skewness

The selected threshold of balance detection sensitivity is with computational efficiency. Lower thresholds risk false negatives (missed faults), while higher thresholds increase false positives (unnecessary YOLO analysis).

F. Integration with YOLO Pipeline

The 22 segments classified as "Potentially Abnormal" proceed to the spectrogram generation stage, where 640×640 -pixel Mel-spectrograms are created using the inferno colormap. These images serve as direct inputs to the YOLO object detection model, which performs visual fault identification through bounding box regression and classification.

This two-stage approach (TOPSIS filtering $\rightarrow$ YOLO detection) leverages the strengths of both methodologies:

- TOPSIS: Rapid statistical screening based on acoustic features.
- YOLO: Deep visual analysis for precise fault localization and classification.

G. Comparative Analysis

Table V compares the proposed TOPSIS-YOLO pipeline with alternative approaches.

The proposed methodology achieves near-equivalent accuracy (93.8% vs. 94.2%) while reducing processing time by 85%, demonstrating superior computational efficiency without sacrificing detection performance.

TABLE V. COMPARISON OF FAULT DETECTION APPROACHES

Approach	Processing Time	Accuracy	Computational Cost
YOLO Only (All Segments)	Baseline (100%)	94.2%	High
SVM Classification	68%	89.5%	Medium
Random Forest	72%	91.3%	Medium
TOPSIS + YOLO (Proposed)	15%	93.8%	Low

Note: Processing time normalized to full YOLO analysis baseline.

H. Limitations and Future Work

While the TOPSIS-based filtering demonstrates strong performance, several limitations warrant consideration:

- Threshold dependency: Optimal threshold values may vary across different machinery types.
- Feature weighting: Equal weights were used; domain-specific weighting could improve discrimination.
- Dataset size: Validation on larger, more diverse datasets is recommended.

Future research directions include adaptive threshold optimization through reinforcement learning and multi-class TOPSIS extensions for fault type differentiation.

VI. CONCLUSION

This study presented a novel hybrid fault detection framework that synergistically combines TOPSIS-based mathematical filtering with YOLO deep learning for efficient audio-based industrial machinery monitoring. The experimental evaluation on 150 industrial audio segments demonstrated that the TOPSIS pre-filter achieved robust discriminative capability, with clear bimodal separation between normal (mean $C_i = 0.847 \pm 0.025$) and abnormal (mean $C_i = 0.084 \pm 0.033$) segments using a classification threshold of $\tau = 0.65$. By successfully identifying 85.3% of segments as normal and eliminating unnecessary processing, the hybrid system maintained high detection accuracy (93.8%) comparable to full-dataset YOLO analysis (94.2%) while reducing processing time by 85%, achieving approximately $6.8\times$ faster performance. The decision fusion logic proved effective in handling edge cases, allowing YOLO to override false positives and flagging low-confidence detections for expert review. Analysis revealed that spectral features (Spectral Centroid, Spectral Bandwidth, and Peak Frequencies) contributed most significantly to fault discrimination, with abnormal segments exhibiting 35-40% higher spectral

centroids, 60-75% broader bandwidths, and positive kurtosis values exceeding 2.5, aligning with established acoustic signatures of mechanical faults. The modular architecture provides practical advantages for industrial deployment, enabling independent optimization of TOPSIS thresholds and YOLO models while maintaining transparency and interpretability essential for safety-critical applications.

While the proposed system demonstrates strong performance, several opportunities exist for future enhancement. The optimal TOPSIS threshold was empirically determined and may require adjustment for different machinery types; domain-specific feature weighting could further improve discrimination; and validation on larger, more diverse datasets is necessary to establish generalizability. Future research directions include adaptive threshold optimization through reinforcement learning, extension to multi-class fault categorization, integration of temporal dependencies through recurrent architectures to capture evolving fault signatures, exploration of alternative multi-criteria decision-making methods (VIKOR, PROMETHEE, fuzzy AHP), and incorporation of transfer learning to accelerate deployment across diverse industrial settings. Beyond immediate technical contributions, this research demonstrates the value of hybrid approaches combining classical mathematical methods with modern deep learning, challenging the assumption that deep learning must be applied uniformly to all data and highlighting opportunities for computational efficiency through intelligent data triage. The proposed TOPSIS-YOLO framework represents a practical, deployment-ready solution that achieves the critical balance between diagnostic accuracy and operational efficiency, providing a blueprint for scalable industrial AI systems. The experimental validation confirms readiness for pilot deployment, with future work focusing on large-scale field trials, multi-site validation, and extension to diverse machinery types to establish this approach as a standard solution for audio-based industrial fault detection.

ACKNOWLEDGMENT

We would like to thank Cebu Technological University for supporting this study.

REFERENCES

- [1] B. A. Tama, M. Vania, S. Lee, and H. Lee, "Recent advances in the application of deep learning for fault diagnosis of rotating machinery using vibration signals," *Artificial Intelligence Review*, vol. 56, pp. 4667–4709, May 2023.
- [2] A. Hämmäläinen, A. Karhinen, J. Miettinen, Z. Dahl, and R. Viitala, "Rapid Deployment of Deep Learning-Based Condition Monitoring for Rotating Machines," *IEEE Access*, vol. 13, pp. 1–1, Jan. 2025, doi: 10.1109/ACCESS.2024.0429000.
- [3] M. Tiboni, C. Remino, R. Bussola, and C. Amici, "A review on vibration-based condition monitoring of rotating machinery," *Applied Sciences*, vol. 12, no. 3, p. 972, 2022.
- [4] X. Ye and G. Li, "An intelligent fault diagnosis method for rolling bearing using motor stator current signals," *Measurement Science and Technology*, vol. 35, no. 8, p. 086131, 2024.
- [5] J. Kim, H. Lee, S. Jeong, and S. H. Ahn, "Sound-based remote real-time multi-device operational monitoring system using a convolutional neural network (CNN)," *Journal of Manufacturing Systems*, vol. 58, pp. 431–441, 2021.
- [6] J. L. Camarena-Martinez et al., "Machine fault detection using acoustic signals and deep learning," *Applied Sciences*, vol. 10, no. 18, p. 6468, 2020.
- [7] S. S. Dlay, M. A. Al-Jumaily, and W. S. Khan, "Feature extraction techniques for machine condition monitoring using sound signals: A review," *Journal of Sound and Vibration*, vol. 482, p. 115441, 2020.
- [8] H. Koizumi, T. Okamoto, and Y. Mitsufuji, "Deep learning-based sound source identification for industrial machine monitoring," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 875–887, 2020.
- [9] F. Alharbi, S. Luo, H. Zhang, K. Shaukat, G. Yang, C. A. Wheeler, and Z. Chen, "A Brief Review of Acoustic and Vibration Signal-Based Fault Detection for Belt Conveyor Idlers Using Machine Learning Models," *Sensors*, vol. 23, no. 4, p. 1902, Feb. 2023.
- [10] M. T. Nguyen and J. H. Huang, "Fault detection in water pumps based on sound analysis using a deep learning technique," *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.*, vol. 236, no. 7, pp. 3667–3677, 2022.
- [11] K. Dohi, T. Nishida, H. Purohit, R. Tanabe, T. Endo, M. Yamamoto, Y. Nikaido, and Y. Kawaguchi, "MIMII DG: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection for Domain Generalization Task," arXiv preprint, May 2022.
- [12] Y. Li, M. Cheng, K. Jiao, X. Shi, S. Li, and C. Li, "A survey of mechanical fault diagnosis based on audio signal analysis," *Measurement*, vol. 220, 113294, Dec. 2023.
- [13] A. J. Kulkarni (Ed.), *Multiple Criteria Decision Making: Techniques, Analysis and Applications*, Springer, 2022.
- [14] M. Hussain, "YOLOv5, YOLOv8 and YOLOv10: The Go-To Detectors for Real-Time Vision," arXiv preprint, Jul. 2024.
- [15] G. Jocher et al., "YOLOv5: A state-of-the-art real-time object detection system," GitHub Repository, 2021. [Online]. Available: <https://github.com/ultralytics/yolov5>
- [16] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv preprint, Apr. 2020.
- [17] C. Li, L. Li, H. Jiang, K. Weng, Y. Geng, L. Li, Z. Ke, Q. Li, M. Cheng, W. Nie, Y. Li, B. Zhang, Y. Liang, L. Zhou, X. Xu, X. Wei, and X. Wei, "YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications," arXiv preprint, Sept. 2022.
- [18] M. L. Ali and Z. Zhang, "The YOLO Framework: A Comprehensive Review of Evolution, Applications, and Benchmarks in Object Detection," *Computers*, vol. 13, no. 12, p. 336, 2024.
- [19] R. Sapkota, M. Karkee, "YOLO Advances to Its Genesis: A Decadal and Comprehensive Review of the You Only Look Once (YOLO) Series," *Artificial Intelligence Review*, 2025.
- [20] J. R. Terven and D. M. Cordova-Esparaza, "A Comprehensive Review of YOLO: From YOLOv1 to YOLOv8 and Beyond," arXiv preprint, Apr. 2023.
- [21] A. Terven and D. Cordova-Esparaza, "A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *Mach. Learn. Knowl. Extr.*, vol. 6, no. 4, pp. 1680–1716, Jan. 2024.
- [22] G. Di Bona, D. Falcone, A. Forcina, L. Silvestri, F. De Carlo, and M. M. Abaci, "A Hybrid Multi-Criteria Decision Model (HMCMDM) based on AHP and TOPSIS analysis to evaluate Maintenance Strategy," in *Proc. 33rd European Modeling & Simulation Symposium (EMSS 2021)*, Cassino, Italy, 2021, pp. 396–406.
- [23] A. J. Kulkarni, *Multiple Criteria Decision Making: Techniques, Analysis and Applications*, Springer, 2022.
- [24] B. Luo, H. Wang, H. Liu, B. Li, and F. Peng, "Early fault detection of machine tools based on deep learning and dynamic identification," *IEEE Trans. Ind. Electron.*, vol. 66, no. 1, pp. 509–518, Jan. 2019.
- [25] O. S. Gidiagba and J.-H. Jang, "Multi-criteria decision support for sustainable supplier evaluation in mining SMEs: A fuzzy logic and TOPSIS approach," *Sustainability*, vol. 17, no. 3, pp. 132, 2025.
- [26] R. Susmaga and I. Szczech, "Utility Inspired Generalizations of TOPSIS," arXiv preprint, Apr. 2025.
- [27] H. Lee and J. Yu, "A Fault Detection Framework for Rotating Machinery with a Spectrogram and Convolutional Autoencoder," *Applied Sciences*, vol. 15, no. 14, p. 7698, Jul. 2025.

- [28] Dipartimento di Ingegneria Meccanica e Aerospaziale, "Deep Transfer Learning for Machine Diagnosis: From Sound and Music Recognition to Bearing Fault Detection," *Applied Sciences*, vol. 11, no. 24, p. 11663, Dec. 2021.
- [29] M. E. Iglesias-Martínez, J. A. Antonino-Daviu, L. Dunai, J. A. Conejero, and P. F. de Córdoba, "Higher-Order Spectral Analysis and Artificial Intelligence for Diagnosing Faults in Electrical Machines: An Overview," *Mathematics*, vol. 12, no. 24, p. 4032, Dec. 2024.
- [30] Z. Zhang, Y. Liu, M. Zhang, and Y. Chen, "Unsupervised Fault Diagnosis for Rotating Machinery Using Wide-Kernel Convolutional Autoencoder With Attention," *Sensors*, vol. 24, no. 24, p. 8053, Dec. 2024.
- [31] R. Susmaga, I. Szczech, and D. Brzeziński, "Towards Explainable TOPSIS: Visual Insights into the Effects of Weights and Aggregations on Rankings," *arXiv preprint*, Jun. 2023.
- [32] B. Xin, X. Zhang, C. Yuan, and C. Li, "Fault Diagnosis of Wind Turbine Based on CNN-LSTM Parallel Network Model," *Academic Journal of Engineering and Technology Science*, vol. 6, no. 12, 2023.
- [33] Z. Wang, H. Meng, J. Liu, and G. Meng, "Optimized YOLOv8 Model for Aerial Pedestrian Detection in Drone-Based Monitoring Systems," in *Proc. 2nd Int. Conf. Machine Learning & Intelligent Computing*, PMLR, 2025, pp. 395–403.
- [34] Y. Liu, Z. Zhao, F. Liang, and L. Bo, "Intelligent Fault Diagnosis for Rotating Machinery Utilizing Gramian Angular Field-Parallel CNN and GRU Networks," *Applied Sciences*, vol. 15, no. 16, p. 9217, 2025.
- [35] H. Zhou, Y. He, J. Luo, and Y. Lin, "A Layered Framework for Universal Extraction and Recognition of Electrical Diagrams," *Electronics*, vol. 14, no. 5, p. 833, Mar. 2025.
- [36] T. Zonta, C. A. da Costa, R. da Rosa Righi, M. J. de Lima, and E. Silveira da Trindade, "Predictive maintenance in the Industry 4.0: A systematic literature review," *Computers & Industrial Engineering*, vol. 150, p. 106889, 2020.
- [37] Z. M. Çınar, A. Abdussalam Nuhu, Q. Zeeshan, O. Korhan, M. Asmael, and B. Safaei, "Machine Learning in Predictive Maintenance Towards Sustainable Smart Manufacturing in Industry 4.0," *Sustainability*, vol. 12, no. 19, p. 8211, 2020.
- [38] O. Kullu and E. Çınar, "A Deep-Learning-Based Multi-Modal Sensor Fusion Approach for Detection of Equipment Faults," *Machines*, vol. 10, no. 11, p. 1105, Nov. 2022.
- [39] M. Gan, C. Wang, and C. Zhu, "Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings," *Mech. Syst. Signal Process.*, vol. 72-73, pp. 92-104, May 2016.
- [40] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84-90, Jun. 2017.
- [41] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: A review and roadmap," *Mech. Syst. Signal Process.*, vol. 138, 106587, Apr. 2020.
- [42] S. Tang, S. Yuan, and Y. Zhu, "Cyclostationary analysis towards fault diagnosis of rotating machinery," *Processes*, vol. 8, no. 10, pp. 1217, Oct. 2020.
- [43] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations (ICLR)*, San Diego, CA, USA, 2015, pp. 1-15.
- [44] N. P. Wandji, H. Trelu, A. Bissuel, N. Tchafa, and O. Hubert, "An interpretable fault detection approach for industrial processes based on improved autoencoder," *IEEE Trans. Ind. Inform.*, vol. 20, no. 7, pp. 9124-9135, Jul. 2024.
- [45] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5990-5998, Jul. 2018.
- [46] A. Payandeh, K. T. Baghaei, P. Fayyazsanavi, S. B. Ramezani, Z. Chen, and S. Rahimi, "Deep Representation Learning: Fundamentals, Technologies, Applications, and Open Challenges," *IEEE Access*, vol. 11, pp. 45624–45649, 2023.
- [47] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Representations (ICLR)*, San Diego, CA, USA, 2015, pp. 1-14.