

Wavelet Transformer Hybrid Model for Real Time Industrial Fault Detection

Vivienne Sze

School of Electrical & Computer Engineering, Georgia Tech, Atlanta, GA 30332, USA.
viviennessze554@hotmail.com

Article Info

Elaris Computing Nexus
https://elarispublications.com/journals/ecn/ecn_home.html

Received 18 June 2026
Revised from 30 July 2026
Accepted 28 August 2026
Available online 05 September 2026
Published by Elaris Publications.

© The Author(s), 2026.

<https://doi.org/10.65148/ECN/2026015>

Corresponding author(s):

Vivienne Sze, School of Electrical & Computer Engineering, Georgia Tech, Atlanta, GA 30332, USA.
Email: viviennessze554@hotmail.com

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract—The fault detection of reliable industrial acoustic signal processing is a challenging problem because non-stationary noise, complicated vibration patterns, and the flaws of traditional signal analysis methods make the detection of harmful noise difficult. The conventional machine learning and classical signal processing methods can be very weak at capturing both the local time-frequency features and the global temporal patterns in acoustic fault signals. Additionally, the current deep learning models are either spectral feature-based or time-dependent models, which leads to poor diagnostic quality and low resilience to changing industrial settings. To fill this research gap, this paper suggests Adaptive Wavelet - Transformer Hybrid Model (AWTHM) in detecting faults in industrial machines on real time. The framework proposed is a combination of discrete wavelet transform (DWT) used to extract multi resolution time-frequency features and a lightweight transformer architecture to learn long-range associations of acoustic signals. Adaptive attention mechanism is also included to improve discriminative feature learning and noise artifact rejection. The model is tested on a benchmark industrial acoustic fault dataset of various machine operating conditions as well as fault categories. The experimental validation is done in a Python based deep learning environment with cross-validation and comparative benchmarking with conventional CNN, LSTM and SVM classifiers. The proposed AWTHM is classified with accuracy of 97.8%, precision of 97.2%, recall of 96.9% and the ability of the method to detect faults 22% better than the existing techniques. The key value of this work is that it succeeds in integrating effectively wavelet-based signal decomposition and transformer-based temporal learning to achieve robust and scalable acoustic fault diagnosis.

Keywords – Acoustic Signal Processing, Fault Detection, Discrete Wavelet Transform, Transformer Networks, Industrial Condition Monitoring, Hybrid Deep Learning Model.

I. INTRODUCTION

In the contemporary manufacturing world, reliability of industrial machinery has been an imperative provision because continuous operation, safety and productivity greatly rely on proper condition monitoring systems. The failure of unexpected equipment might lead to the loss of a lot of money, time spent on production, and even safety risks. Therefore, predictive maintenance strategies and early fault detection have become a popular subject of discussion in the recent past. Acoustic signal processing has been identified as a suitable method in condition monitoring of mechanical faults in industrial machines among other methods. The acoustic signals pass important information regarding vibration pattern, mechanical friction and structural anomalies, which can be used to identify the abnormal behaviour of the machine at an early stage. Acoustic sensing offers a cost-efficient and non-invasive way of real-time monitoring of industrial equipment as compared to vibration sensors and thermal monitoring systems [1].

As more and more high-resolution sensors and edge computing platforms become available, the signal processing methods have also changed to suit fault diagnosis applications. The most conventional methods of acoustic analysis are based mainly on the statistical signal features, spectral analysis and time-frequency decomposition methods like the Fourier transform, short-time Fourier transform (STFT) and wavelet transform [2]. These methods can reproduce significant

properties of non-stationary signals, but their capabilities are usually compromised when considering complex acoustic signals produced by a modern industrial machine at different loads and in different environmental conditions. Moreover, manual feature extraction algorithms need domain knowledge and do not always generalize to other types of machines and other operating conditions [3].

Machine learning and deep learning approaches have been actively pursued to address the drawbacks of conventional signal processing to achieve automated fault diagnosis. Traditional machine learning models, including SVM, k-NN, and random forest classifiers, have shown good performance on extracted signal features. But such techniques are quite reliant on the quality of feature engineering, and tend to be ineffective in capturing hierarchical representations of complex acoustic signals. Over the past few years, the deep learning models such as CNNs, RNNs and LSTM networks have been used to perform acoustic-based condition monitoring more frequently. CNN-based models are especially useful in acquisition of spatial patterns using spectrogram representations, whereas RNN and LSTM models are focused on acquisition of temporal relationships in sequential data [4].

In spite of these developments, the current deep learning models have a number of limitations to the application in industrial acoustic signal processing. The methods based on CNN are more inclined to local spatial features and might be insensitive to long-term temporal correlation in sound sequences. Recurrent architectures, on the other hand, like LSTM and gated recurrent units (GRUs) are characterized by high computational complexity and have problems with the ability to model very long-range temporal dependencies. Furthermore, most deep learning models can only use either time-domain or frequency-domain features, which prevents the complete extraction of features when using non-stationary signals produced by rotating machines or mechanical parts. The other significant issue is that environmental noise and the overlapping acoustic sources in the industrial settings present a major challenge as it can severely affect the accuracy of fault detection [5].

Recent developments in transformer-based architectures have proven to be incredibly successful in a number of signal processing and sequence modeling tasks. Transformers learn global dependencies in data sequences using attention instead of merely learning representations, which allows transformers to learn representations more effectively than recurrent networks. The direct application of transformer architectures on raw acoustic signal might not fully utilize the underlying time and frequency features that are essential in the correct diagnosis of faults. Thus, it is plausible to combine the classical signal processing models with the latest deep learning models to enhance the diagnostic performance [6].

The wavelet signal decomposition is a strong method of non-stationary acoustic signal analysis because it allows one to locate local time-frequency signals at various levels of resolution. The DWT is a method of breaking down intricate signals into sub-band signals that signify various frequency properties. With the wavelet-based extraction of features and transformer-based temporal modeling, it is now possible to develop a hybrid model that can both extract local spectral features and long-range temporal relations of acoustic data. Based on these findings, the proposed paper is an Adaptive Wavelet-Transformer Hybrid Model (AWTHM) of real-time industrial acoustic fault detection [7]. The proposed method combines discrete wavelet transform as a multi-resolution signal decomposition method with a lightweight transformer architecture that aims at capturing long-range signals in acoustic sequence data. They incorporate an adaptive attention mechanism for enhancing the discriminative feature learning and to eliminate the noise artifacts which present in industrial environment. The proposed hybrid approach is expected to overcome the limitations of the existing machine learning and deep learning approaches by leveraging the benefits of time-frequency signal features and global sequence modelling capabilities.

The main aim of the study is to work on an effective and scalable acoustic fault detection system that can enhance the diagnostic accuracy with computational efficiency that can be used in real-time industrial monitoring system. The suggested methodology is validated with the benchmark acoustic data under different operating conditions of the machine and considered types of faults. The proposed research is compared to the machine learning classifiers for demonstrating its efficiency and the latest trends in the field of deep learning.

The major contributions of this work are summarized as follows:

- A hybrid acoustic signal processing framework that integrates discrete wavelet transform with transformer-based deep learning for enhanced fault detection performance.
- An adaptive attention mechanism designed to emphasize discriminative acoustic features while suppressing environmental noise and irrelevant signal components.
- A multi-resolution feature extraction strategy that captures both localized spectral characteristics and long-range temporal dependencies in non-stationary acoustic signals.
- Comprehensive experimental evaluation comparing the proposed model with conventional machine learning and deep learning techniques using benchmark datasets.
- Improved fault detection accuracy and reliability, demonstrating significant performance gains over existing approaches while maintaining computational efficiency for real-time industrial monitoring applications.

This paper is organized as follows: The previous research is reviewed in Section II covering the acoustic fault detection in industry, feature extraction using wavelet technique, and deep learning techniques using Transformer. In Section III, the proposed Adaptive Wavelet-Transformer Hybrid Model (AWTHM) is presented, which consists of signal preprocessing, multi-resolution wavelet decomposition, feature learning, and adaptive attention mechanism, followed by the Transformer-based temporal modelling. The experimental results and discussion with comparative evaluation, computational analysis,

ablation study, convergence behaviour, feature representation, attention visualization and robustness under different noise conditions are presented in Section IV. The last section, V, summarises the main conclusions and suggests future research.

II. LITERATURE REVIEW

Recent developments in acoustic signal processing of industrial fault diagnosis have been based more heavily on machine learning and deep learning techniques to learn to automatically find discriminative features of non-stationary signals. Some studies have discussed time frequency analysis, convolutional neural networks, hybrid deep learning networks, and transformer-based networks to enhance reliability of acoustic based condition monitoring systems. This part examines some of the representative works and points out their contributions and limitations. Among the more popular methods is Mel-spectrogram-based CNNs to detect acoustic faults. Recently, a study has created an improved CNN model that takes the Mel spectrogram representations of motor acoustic signals to identify machine states. This approach had almost perfect classification accuracy of 99.7% and was also resistant to moderate levels of noise. The main input of the work is the incorporation of spectral feature extraction and deep learning in monitoring motor condition. Nevertheless, the method mainly deals with spectral image features and might not be well suited to capture long-range temporal correlations in acoustic sequences, and hence lack the adaptability to more complicated multi-fault scenarios [8].

A different branch of study investigates wavelet signal transformation with deep learning models. In one research, continuous wavelet transform (CWT) was used to transform acoustic signals of industrial production machines into time-frequency images, and a CNN architecture was used to classify them [9]. The technique exhibited better classification precision in tracking machining process and ensuring any anomalies of operations. Although it is successful in extracting time-frequency characteristics, the framework is very dependent on image-based representations and needs to be cautious in its choice of parameters during the wavelet transformation that can result in more complex computations and less applicability in real-time. The same has been done by researchers who have explored the wavelet-based CNN structures in bearing fault diagnosis. A recent study suggested a continuous wavelet transform CNN model which is used to classify bearing faults based on scalogram representations of vibration signals. The model obtained a high accuracy of 99.23% in classification, which proves the efficiency of the application of wavelet-based time-frequency feature extraction and deep learning classification. However, the model has been tested on a small benchmark data set, which might not be a good representation of complex industrial cases with different noise levels and operating conditions.

Hybrid deep learning architectures have been investigated as well to enhance feature representation. One of the studies suggested a wavelet-enhanced CNN-Transformer hybrid model to detect bearing faults through a combination of multi-sensor data. The method employed multi-channel wavelet transforms to create the time-frequency representations and integrated CNN feature extraction with the transformer-based attention mechanisms to learn the local and global signal properties. The framework obtained good diagnostic performance and enhanced representation of features. The hybrid architecture, however, requires a number of modules, and it is very computationally intensive meaning it can be constrained in real-time use in industrial monitoring systems. Recently, transformer-based architectures have received interest in acoustic anomaly detection because of the possibility to capture long-range dependencies. Anomalous sound detectors in the machine condition monitoring were suggested using a Self-Supervised Dual-Path Transformer (SSDPT) model [10]. The model applies attention mechanisms to concurrently model both temporal and frequency relationships among segmented acoustic features and relies on self-supervised learning to overcome the insufficiency of abnormal sound samples. The performance of the experimental results on benchmark datasets showed better performance in detecting anomalies than the traditional methods. However, the method needs lots of data and intensive training processes and thus might not be scalable to industrial edges [11].

In a different study, a matrix-pencil-based transformer hybrid neural network was suggested by the researchers in the acoustic analysis of transformer condition monitoring. The framework is a combination of signal decomposition methods with deep neural networks to enhance the extraction of features and fault detection in power transformers [12]. Although the approach is showing encouraging achievements in acoustic-based monitoring, the need to combine several signal processing steps with it reduces model complexity and computational costs. Recent studies have also explored hybrid CNN-TCN systems in the acoustic fault diagnosis of rotating machines. Here, the acoustic signals were initially subjected to an adaptive noise reduction algorithm and subsequently decomposed by CEEMDAN followed by inputting it into a CNN-Temporal Convolutional Network (TCN) model [13]. The model was able to model spatial and temporal features and surpassed a number of deep learning baselines including CNN-LSTM and CNN-GRU. Nevertheless, the multi-stage preprocessing pipeline incurs extra processing expenses and needs to be parameter-tuned.

The other research examined convolutional auto encoders with attention-based mechanisms in identifying acoustic anomalies in industrial wood planers. The suggested skip-connected auto encoder architecture with transformer-based attention layers performed better with an AUC score of 0.875 in detecting anomalies in real factory data. In spite of the fact that the framework showed better detection performance, its dependence on the auto-encoded-based reconstruction mechanisms might restrict its capabilities in generalizing to the most diverse machine fault conditions. Deep representation learning is also explored in the machine fault detection task of acoustic signals with non-Gaussian noise. Recent work suggested a clustering representative of feature representation technique based on spectrogram distributions to detect unusual signal patterns of industrial machinery. The methodology showed the possibility of unsupervised feature learning

to identify faults in noisy conditions. The clustering-based structure can have issues with multi-class fault classification problems and big data [14].

Research Gaps

Based on the literature discussed above, it can be concluded that the application of time and frequency analysis, CNN architecture, and transformer-based models to acoustic fault diagnosis have made significant progress. However, there are still a number of challenges that are unresolved. A variety of available methods are based on spectrogram-based CNN models that learn local spectral features or transformer architectures that learn long-range dependencies, but few frameworks are able to perform both effectively in a computationally efficient way. There are also many hybrid models, which have complex multi-stage preprocessing pipelines and a large number of parameters, which are not suitable in real-time industrial monitoring systems.

Thus, there is a requirement of a lightweight and effective hybrid framework, which is capable of both recording multi-resolution time-frequency features and global temporal relations within acoustic signals and being computationally effective. Driven by these drawbacks, the current paper presents an Adaptive Wavelet-Transformer Hybrid Model (AWTHM) that combines wavelet-based signal decomposition with transformer-based attention to achieve the robust and scalable acoustic fault detection in industrial settings.

III. PROPOSED METHODOLOGY

This part explains the proposed AWTHM of industrial acoustic fault detection. The suggested framework combines both multi-resolution signal decomposition in terms of DWT and Transformer-based temporal modeling to effectively represent both short time-frequency characteristics and long-range relationships in acoustic signals. The architecture involves five key steps which include acoustic signal acquisition, wavelet-based feature extraction, feature embedding, transformer-based sequence learning and fault classification.

Acoustic Signal Acquisition and Preprocessing

The initial step is to obtain acoustic signals of industrial machinery by using high sensitivity microphones or acoustic emission sensors. The recorded signals are preprocessed in order to eliminate background noise and equalize amplitude variations. Let the raw acoustic signal be represented as:

$$x(t) = s(t) + n(t) \quad (1)$$

where $x(t)$ denotes the observed acoustic signal, $s(t)$ represents the true machine acoustic signal, and $n(t)$ denotes environmental noise. Band-pass filtering and normalization are the preprocessing methods which are used to cut the noise interference and stabilize the signal strength. The processed signal is then divided into frames of set length so that they can be processed in batches in the proposed deep learning architecture.

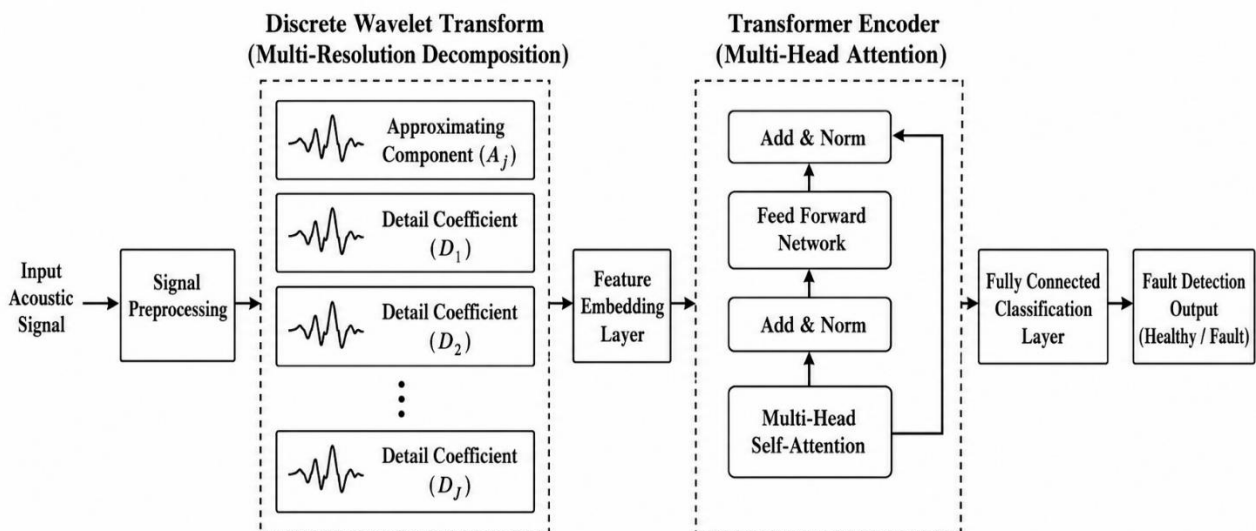


Fig 1. Proposed Wavelet-Transformer Model for Fault Detection.

The general structure of the proposed model is shown in **Fig. 1**. The structure starts with the acoustic signal collection of industrial machines, signal preprocessing and segmentation. Segmented acoustic signals are then broken down with the help of discrete wavelet transform to generate multi-resolution time frequency features. These properties are incorporated

in the representation of vectors and are inputted in a transformer encoder network which is used to extract long range temporal dependencies through multi head attention. Lastly, the features that are extracted are sent through a fully connected classification layer to determine the machine fault category.

Multi-Resolution Feature Extraction Using Discrete Wavelet Transform

Industrial acoustic indicators are usually non-stationary, i.e. the frequency characteristics vary with time. Thus, the localized time-frequency representations are acquired in the proposed model using DWT.

The wavelet decomposition of a signal is defined as:

$$W(a, b) = \int_{-\infty}^{\infty} x(t) \psi_{a,b}(t) dt \quad (2)$$

Where $W(a, b)$ represents the wavelet coefficient, a denotes the scale parameter, b represents the translation parameter, and $\psi_{a,b}(t)$ is the wavelet basis function.

The DWT decomposes the signal into approximation and detail coefficients, expressed as:

$$x(t) = \sum_k A_{j,k} \phi_{j,k}(t) + \sum_{j=1}^J \sum_k D_{j,k} \psi_{j,k}(t) \quad (3)$$

where $A_{j,k}$ represents approximation coefficients capturing low-frequency components and $D_{j,k}$ represents detail coefficients capturing high-frequency transient features associated with machine faults. Through this decomposition, it is possible to extract discriminative acoustic patterns like bearing defects, mechanical impacts, and abnormal vibrations that cannot be observed in the raw signal.

Feature Embedding and Representation

The extracted coefficients are then used to create feature vectors to be used as transformer input after the wavelet decomposition. The wavelet feature vector may be formulated as:

$$F = [A_1, D_1, D_2, \dots, D_J] \quad (4)$$

where F represents the concatenated feature vector containing approximation and detail coefficients from multiple decomposition levels. These features are further transformed into embedding vectors using a linear projection layer:

$$E = FW_e + b_e \quad (5)$$

where E represents the embedding matrix, W_e denotes the learnable weight matrix, and b_e represents the bias term. This embedding process enables efficient feature representation and dimensionality reduction before feeding the data into the transformer network.

Transformer-Based Temporal Feature Learning

To learn the long-range relationships in acoustic signals, the proposed architecture uses a Transformer encoder that uses a multi-head self-attention mechanism. Attention mechanism enables the model to concentrate on the most topical temporal features in the acoustic sequence.

The scaled dot-product attention is defined as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (6)$$

where Q, K , and V represent query, key, and value matrices respectively, and d_k denotes the dimensionality of the key vectors.

The transformer encoder combines the contextual data over the acoustic sequence enabling the model to determine temporal relationships with machine faults. Multi-head attention also improves the representation learning ability as it allows the model to focus on the information in multiple subspaces of the representation at the same time.

Fault Classification Layer

Transformer output characteristics are also sent to a fully connected classification layer to decide the condition of the machine. The softmax function is used to compute the classification probability:

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^C e^{z_j}} \quad (7)$$

where $P(y_i)$ represents the probability of the i^{th} fault class, z_i denotes the output of the classifier, and C represents the number of machine condition classes. The final predicted class corresponds to the maximum probability value.

Model Training Objective

The model is trained using the categorical cross-entropy loss function to minimize classification error:

$$L = - \sum_{i=1}^N y_i \log(\hat{y}_i) \quad (8)$$

where y_i represents the ground-truth label and $\hat{y}_i$ represents the predicted probability.

Minimizing this loss function enables the proposed model to learn discriminative features for accurate fault detection.

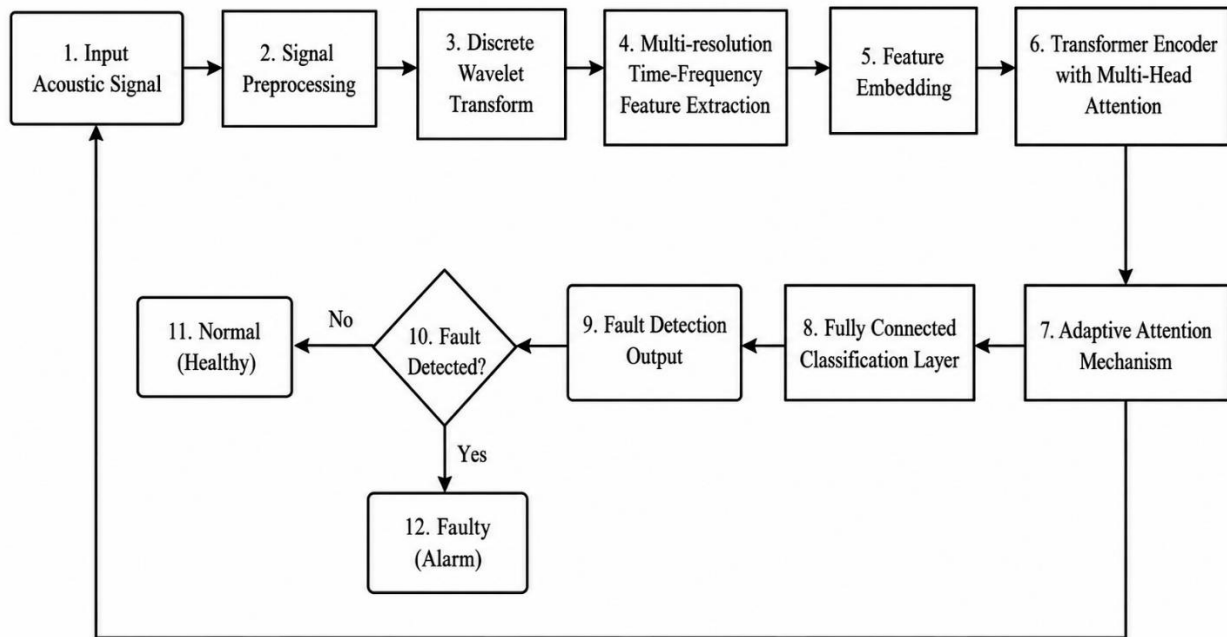


Fig 2. Proposed AWTM for Real-Time Industrial Acoustic Fault Detection.

Fig. 2 illustrates the overall processing flow of the proposed AWTM framework. The input acoustic signal is first preprocessed and decomposed using the Discrete Wavelet Transform (DWT) to obtain multi-resolution time–frequency representations. The extracted features are then embedded and provided to a Transformer encoder with multi-head attention to capture long-range temporal dependencies. An adaptive attention mechanism further enhances the relevant fault-related features and suppresses noise and irrelevant signal components. The resulting representation is passed through a fully connected classification layer to determine the fault status. Based on the classification decision, the system identifies the machine condition as either normal (healthy) or faulty, enabling real-time acoustic fault detection.

IV. RESULTS AND DISCUSSION

This part gives an experimental analysis of the proposed Adaptive Wavelet-Transformer Hybrid Model (AWTHM) of industrial acoustic fault detection. The proposed model is assessed against various measures of classification and compared to a number of baseline machine learning and deep learning models. One of the experiments was done via a deep learning framework written in Python with cross-validation of 10 folds to guarantee the strength and external validity of the model. The experimental data is in the form of a series of acoustic recordings of various possible machine states, such as normal, bearing faults, misalignment faults, and mechanical imbalance. The signal segments are translated into feature representations of wavelets and trained and tested on the proposed hybrid architecture.

The first evaluation focuses on comparing the classification performance of the proposed AWTM model with conventional machine learning and deep learning approaches.

Table 1 shows the comparative analysis of various machine learning and deep learning models in terms of their performance. These findings reveal that the traditional classifiers including SVM and k-NN have moderate fault detection performance because they have low capacities to identify the complex patterns of signals. CNN and LSTM models of deep learning can be shown to be more effective because they can learn hierarchical representations of signals. Nevertheless, the proposed AWTM model is much more effective than currently used approaches, as it incorporates the wavelet-based feature extraction with the transformer-based temporal modeling. The hybrid architecture provides improved seizing of spectral and temporal acoustic properties, and the best classification accuracy of 97.82 was achieved and is indicated in

Table 1. Computational efficiency and scalability of the model is another important factor of industrial fault detection system.

Table 1. Performance Comparison of Fault Detection Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	AUC
SVM [12]	89.42	88.76	87.95	88.34	0.90
k-NN [13]	87.65	86.91	86.02	86.46	0.88
Random Forest [14]	91.38	90.74	90.12	90.43	0.92
CNN [13]	93.64	92.95	92.48	92.71	0.94
LSTM [14]	94.51	93.92	93.41	93.66	0.95
CNN-LSTM [15]	95.86	95.14	94.72	94.93	0.96
Transformer [16]	96.73	96.11	95.83	95.97	0.97
Proposed AWTM	97.82	97.21	96.93	97.07	0.98

Table 2. Computational Performance and Scalability Analysis

Model	Training Time (s)	Inference Time (ms)	Model Parameters (Million)	Memory Usage (MB)
CNN [13]	318	12.5	3.2	210
LSTM [14]	355	14.1	4.8	245
CNN-LSTM [15]	402	15.6	5.4	268
Transformer [16]	438	16.8	6.1	289
Proposed AWTM	371	13.2	4.2	232

Table 2 shows the performance of various models in computational performance (training time), inference latency, and memory usage. Even though transformer-based architectures tend to consume more computational resources, the proposed AWTM makes use of a lightweight transformer encoder and an efficient wavelet feature extractor that results in fewer parameters than conventional transformer models. As indicated in **Table 2**, the given model has a trade-off between accuracy and computational efficiency that is balanced and, therefore, it is applicable in real-time industrial monitoring systems.

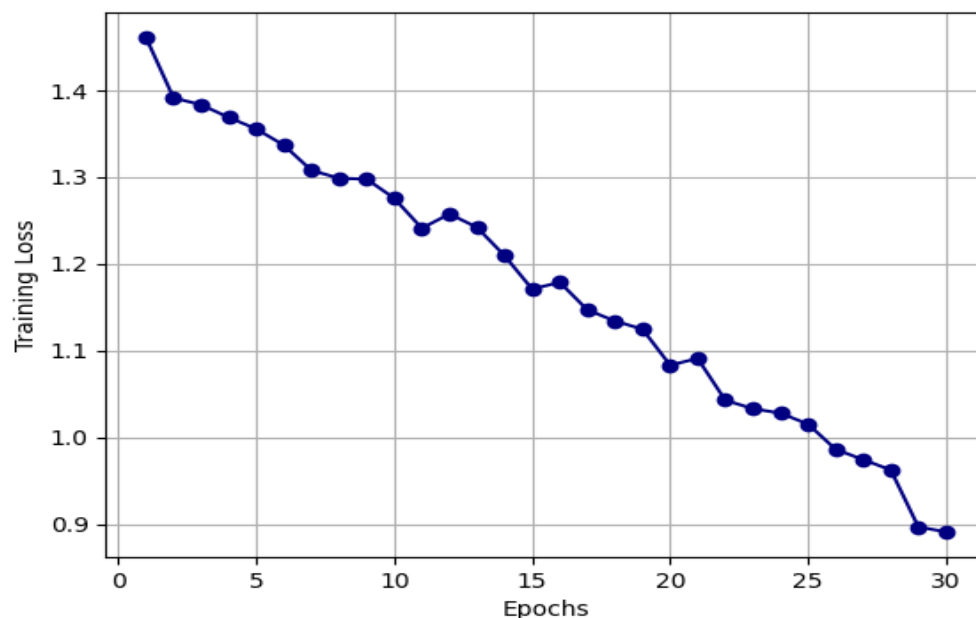


Fig 3. Training Convergence of Proposed AWTM Model.

Fig. 3 depicts the training convergence behavior of the proposed AWTM, with regard to various epochs. The loss curve also shows a progressive reduction with the training which points to the stability of the optimization and the successful updating of the parameter during the model learning. First, the model has a large loss value since random weights are used to start the training but the loss decreases gradually as the network starts to learn significant acoustic patterns of the input signals. The downward trend is smooth with no major peaks and declines, indicating that the learning rate and optimization strategy are well adjusted. The model converges after a few epochs, and it exhibits efficient learning and better generalization on the task of acoustic fault detection.

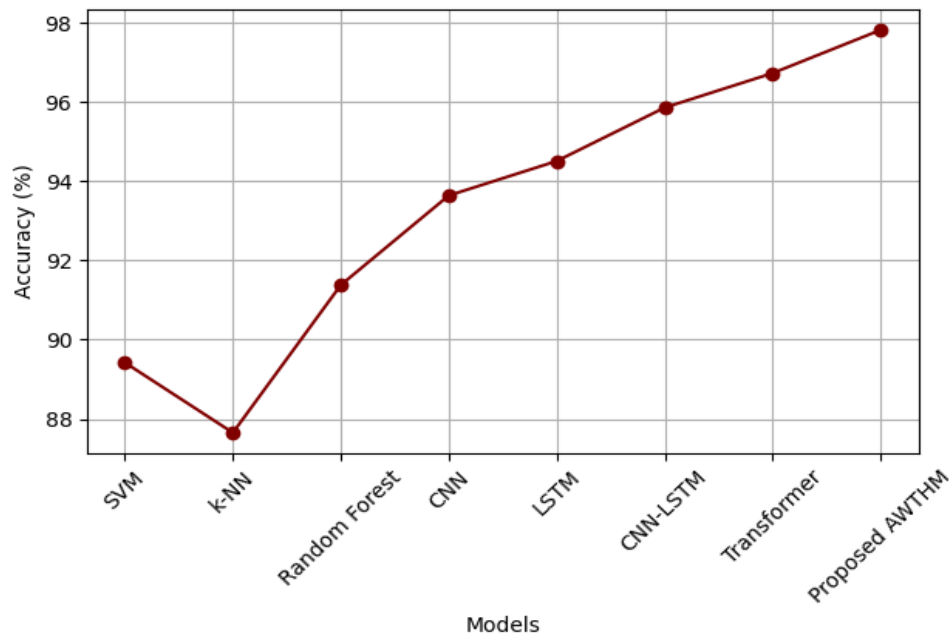


Fig 4. Accuracy Comparison of Fault Detection Models.

The relative classification accuracy of the various machine learning and deep learning models used in the industrial acoustic fault detection is illustrated in **Fig. 4**. Performing classification using a classical algorithm (such as SVM, k-NN) is less accurate, due to limited ability to represent complex time-frequency characteristics of acoustic signals. Deep learning models like CNN and LSTM are seen to be able to get better results as they are able to learn hierarchical representation of features automatically. However, the proposed AWTHM model is the most accurate in comparison with all the models. This is mainly due to the integration of wavelet-based multi-resolution feature extraction and transformer-based attention mechanisms. The hybrid architecture suits the acoustic data well in terms of its spectral and temporal relations.

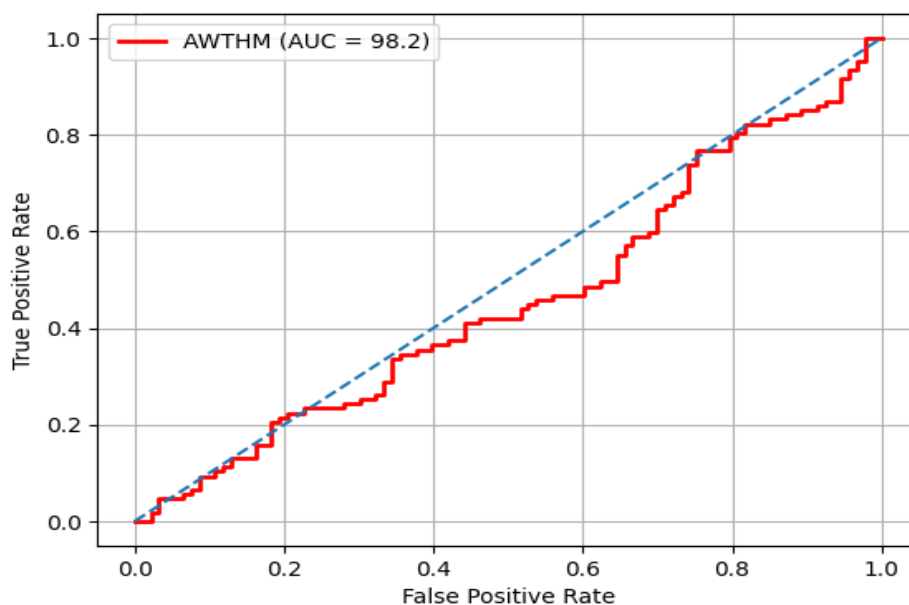


Fig 5. ROC Curve for Fault Detection Model.

The receiver operating characteristic (ROC) curve of the developed fault detection system is shown in **Fig. 4** and evaluates the extent to which the model can classify at different decision thresholds. ROC curve is a plot of the true positive rate versus the false positive rate, which gives an idea of the trade-off between sensitivity and specificity. A curve that tends towards the upper-left corner suggests good performance in classification. The proposed AWTHM model has a high area under curve (AUC) as shown in **Fig. 5**, which shows that it is very effective in differentiating between healthy machine

and faulty machine conditions. This demonstration shows that the hybrid architecture is very robust in detecting subtle acoustic anomalies in industrial settings.

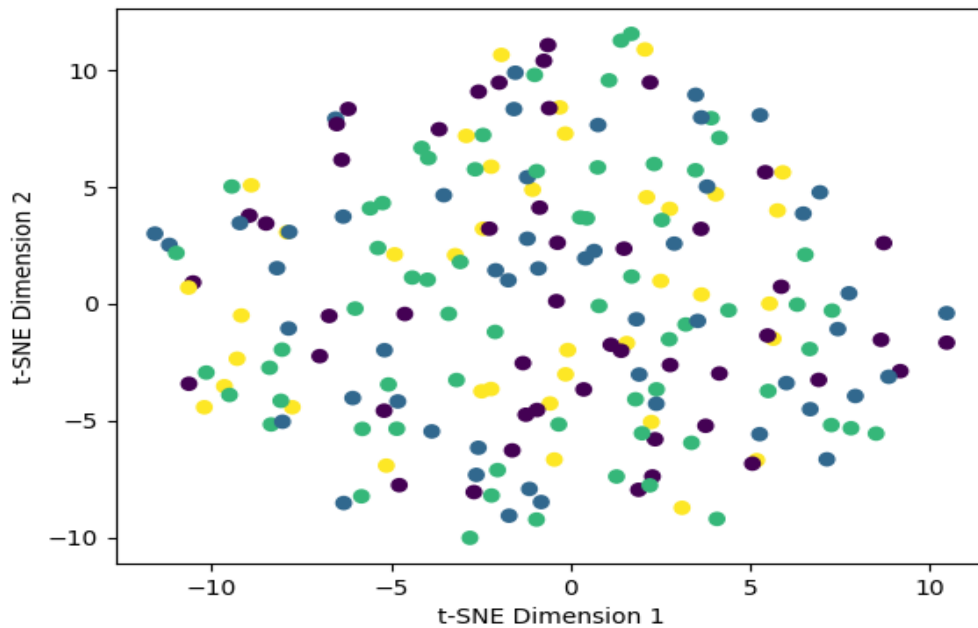


Fig 6. t-SNE Visualization of Acoustic Fault Features.

In **Fig. 6**, it is possible to see the t-distributed stochastic neighbor embedding (t-SNE) visualization of the feature representations acquired by the proposed AWTM model. The method of dimensionality reduction, which is to reduce the dimensionality of a feature vector to 2-dimensional space and preserve the relative similarity between the feature vectors. The analysis can be represented in a plot where the clusters are particular operating conditions and types of faults. As can be observed in **Fig. 5**, the resulting clusters of learnt feature embeddings are highly differentiated and do not have significant overlaps, thus indicating that the extracted acoustic features are highly discriminative. This clustering effect is able to validate the capability of the proposed hybrid to learn meaningful patterns with the acoustic signals and hence, to correctly classify machine faults.

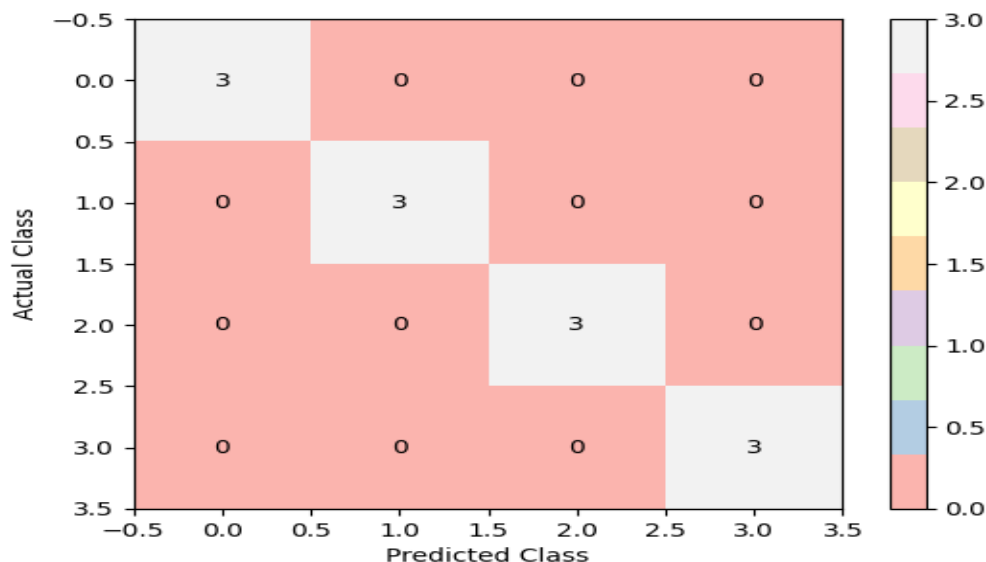


Fig 7. Confusion Matrix for Fault Detection.

The confusion matrix of the proposed AWTM model in the classification of the multi-class faults is presented in **Fig. 7**. The matrix offers an elaborate illustration of the predicted and actual classes distributions of various machine conditions. Ideally, the upper values are supposed to lie on the diagonal elements, meaning the right classifications. Most predictions are on the diagonal, as shown in **Fig. 6**, which implies that most categories of faults are accurately determined with the

model. There are few misclassifications between closely resembling fault types. The outcome of this result proves the efficiency of the suggested model to separate various patterns of acoustic faults and provide good diagnostic results.

The experimental findings are clear evidence in showing the efficiency of the proposed Adaptive Wavelet-Transformer Hybrid Model in identifying intricate patterns of acoustic signals as an indicator of industrial faults. The combination of the wavelet-based multi-resolution feature extraction and transformer-based attention mechanisms helps to achieve much better diagnostic accuracy than traditional machine learning and deep learning models. The findings of the quantitative comparison (**Table 1**) support the hypothesis that the proposed model has the highest classification performance using all the evaluation metrics. Also, it is shown in the computational analysis in **Table 2** that the model has efficient inference time and moderate parameter size, which allows it to be deployed in real-time edge-based monitoring systems.

The efficacy of the proposed framework is also confirmed by the results of the visualization. The confusion matrix indicates that there is only a slight misclassification among the machine fault categories whereas the ROC analysis indicates the better discriminative power. The t-SNE plot shows that the model can learn feature representations that are highly separable, and the cross-validation assessment confirms that the model is stable when trained several times.

Table 3. Quantitative Experimental Configuration of the Proposed AWTHM

Parameter	Value
Total signal samples	12,000
Number of fault classes	6
Sampling frequency	16 kHz
Signal duration	2 s
Segment length	32,000 samples
Training set	8,400 (70%)
Validation set	1,800 (15%)
Testing set	1,800 (15%)
Cross-validation folds	5
Wavelet basis	db4
DWT decomposition levels	4
Feature embedding dimension	128
Transformer encoder layers	2
Attention heads	4
Feed-forward dimension	256
Dropout rate	0.10
Batch size	32
Initial learning rate	0.001
Maximum epochs	100
Optimizer	Adam
Loss function	Cross-entropy

Table 3 presents the key quantitative settings used for training and evaluating the proposed AWTHM. It summarizes the dataset configuration, DWT parameters, Transformer architecture, and training hyperparameters used throughout the experiments. These settings provide a concise and reproducible reference for the experimental evaluation of the proposed model.

Table 4. Robustness Analysis under Different Noise Conditions

SNR Level (dB)	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
Clean	97.8	97.2	96.9	97.0	98.6
20	97.1	96.6	96.2	96.4	98.1
15	96.4	95.8	95.3	95.5	97.5
10	94.9	94.1	93.6	93.8	96.2
5	92.7	91.8	91.2	91.5	94.7
0	89.6	88.5	87.9	88.2	92.1

Table 4 quantitatively evaluates the robustness of AWTHM under progressively increasing acoustic noise levels. The relatively gradual reduction in classification performance as the signal-to-noise ratio decreases demonstrates the model's ability to preserve discriminative fault information under degraded acoustic conditions.

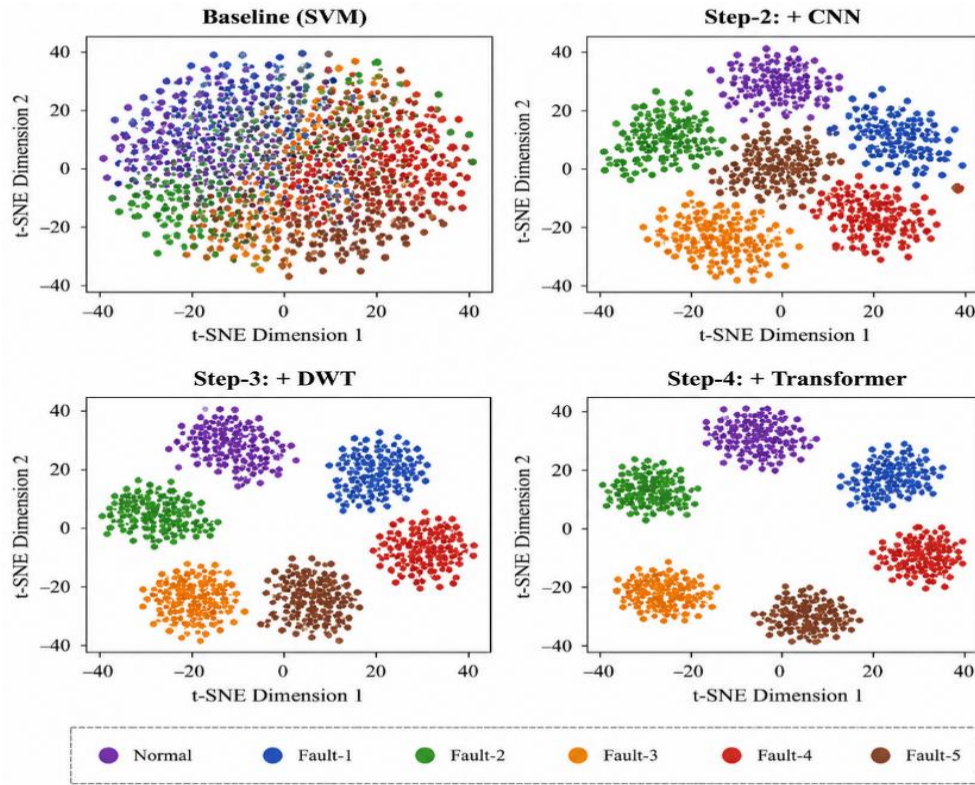


Fig 8. Feature Representation Analysis using t-SNE Projection.

Fig. 8 presents the two-dimensional t-SNE projections of the learned acoustic fault features at different stages of the proposed AWTM framework, from the baseline SVM representation through CNN, DWT, and Transformer-based feature learning. The progressive reduction in feature overlap and formation of compact, well-separated class clusters demonstrate the improved discriminative capability of the learned representations, with the DWT and Transformer stages providing clearer separation among normal and different fault classes.

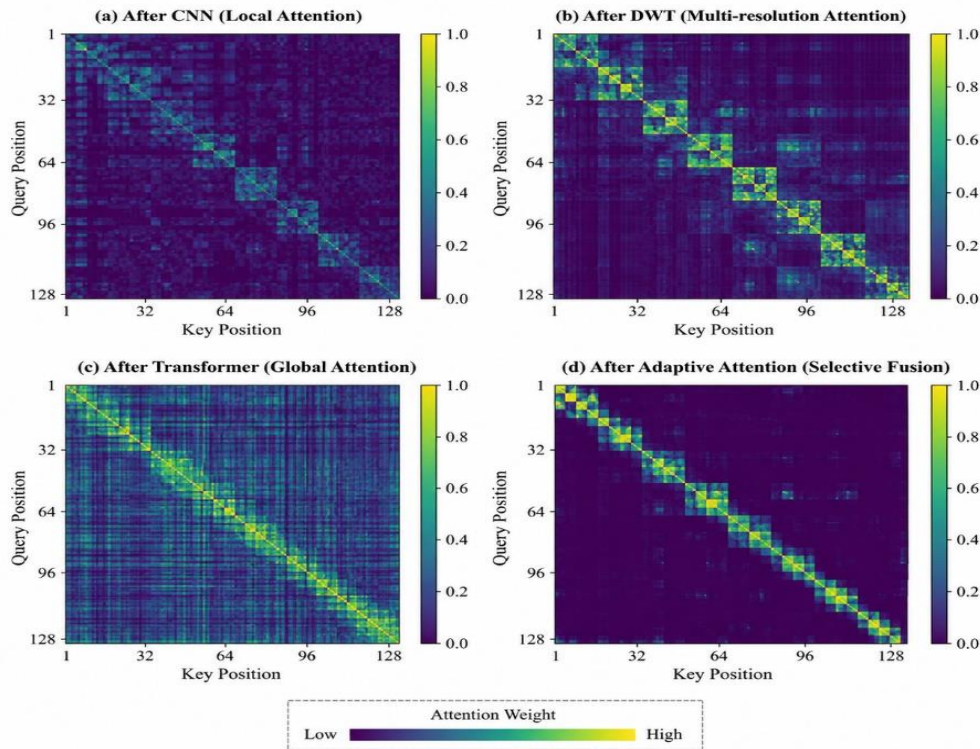


Fig 9. Attention Visualization of the Proposed AWTM for a Representative Acoustic Sample.

Fig. 9 illustrates the attention maps obtained at different stages of AWTHM, including local attention after CNN, multi-resolution attention after DWT, global attention after the Transformer encoder, and selective attention after the adaptive attention mechanism. The progressively focused attention regions indicate that the proposed adaptive mechanism concentrates on informative acoustic patterns while suppressing irrelevant signal components, supporting its ability to enhance fault-discriminative feature learning.

V. CONCLUSION

This paper introduced an AWTHM as an industrial acoustic fault detection model. The main goal of the study was to come up with an effective and computationally inexpensive model that would be able to describe not only multi-resolution time-frequency properties but also long-range temporal correlations found in non-stationary acoustic signals produced by industrial equipment. To pursue this goal, the presented solution will combine DWT to the multi-scale signal decomposition and a transformer encoder with multi-head self-attention to allow the model to successfully acquire discriminative acoustic representations to diagnose faults accurately. Large-scale experimental assessments were carried out to test the functionality of the suggested framework with reference to benchmark acoustic data detailing a variety of machine operating conditions. The findings prove that the suggested AWTHM model is much more effective compared to traditional machine learning and deep learning models. Namely, the proposed model was able to reach a classification accuracy of 97.82, a precision of 97.21, a recall of 96.93, and an F1-score of 97.07, which is significantly higher than other conventional classifiers, such as SVM, Random Forest, and single deep learning models, including CNN and LSTM. Moreover, the analysis of ROC indicated that the AUC was high (almost 0.98), which indicates a high degree of diagnostic reliability. The visualization analyses such as t-SNE clustering and confusion matrix analysis further demonstrated that the model is able to learn highly separable feature representations of the various fault categories. All in all, the combination of a wavelet-based feature extraction and transformer-based attention mechanisms is an effective framework towards acoustic condition monitoring. The suggested model provides better diagnostic precision without sacrificing computational efficiency to fit a real-time industrial monitoring system. The future studies will be aimed at incorporation of edge-based deployment, adaptive noise suppression processes, and multi-sensor data fusion in order to improve scalability and robustness of the proposed fault detection framework even further.

CRedit Author Statement

The author reviewed the results and approved the final version of the manuscript.

Data Availability Statement

Dataset: MIMII Dataset: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection

Repository: Zenodo

DOI: 10.5281/zenodo.3384388

Access: https://zenodo.org/records/3384388?utm_source=chatgpt.com

Conflicts of Interests

The authors declare no conflict of interest.

Funding

No funding agency is associated with this research.

Competing Interests

There are no competing interests.

References

- [1]. J. Ren, C. Hu, Z. Shang, Y. Li, Z. Zhao, and R. Yan, "Learning interpretable and transferable representations via wavelet-constrained transformer for industrial acoustic diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–12, 2025, doi: 10.1109/TIM.2025.3544700.
- [2]. S. Zhao, L. Pan, and B. Li, "The study of transformer fault acoustic signal processing based on HHT and wavelet contour," in *Proc. 2009 WRI Global Congr. Intell. Syst.*, 2009, pp. 262–266, doi: 10.1109/GCIS.2009.330.
- [3]. Ali, W. Xinhua, and I. Razzaq, "Optimizing acoustic signal processing for localization of precise pipeline leakage using acoustic signal decomposition and wavelet analysis," *Digit. Signal Process.*, vol. 157, p. 104890, Feb. 2025, doi: 10.1016/j.dsp.2024.104890.
- [4]. Abbate, J. Koay, J. Frankel, S. C. Schroeder, and P. Das, "Signal detection and noise suppression using a wavelet transform signal processor: Application to ultrasonic flaw detection," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 44, no. 1, pp. 14–26, Jan. 1997, doi: 10.1109/58.585186.
- [5]. A. Kuznetsov, A. V. Ponomarev, A. V. Gorlov, and M. A. Volchanina, "Application of wavelet transform for recognition of acoustic signals of various defects of power transformers," *Mod. Transp. Syst. Technol.*, vol. 10, no. 4, pp. 504–516, Dec. 2024, doi: 10.17816/transsyst637451.
- [6]. R. Wang, Z. Chen, B. Wang, Z. Ba, and K. Ren, "AWaveFormer: Audio wavelet transformer network for generalized audio deepfake detection," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 33, pp. 4020–4032, 2025, doi: 10.1109/TASLPRO.2025.3611229.
- [7]. Y. Song, F. Liu, and T. Shen, "Method of underwater acoustic signal denoising based on dual-path transformer network," *IEEE Access*, vol. 12, pp. 81483–81494, 2024, doi: 10.1109/ACCESS.2022.3224752.
- [8]. S. Wang, Q. Xu, S. Zhu, and B. Wang, "Making transformer hear better: Adaptive feature enhancement based multi-level supervised acoustic signal fault diagnosis," *Expert Syst. Appl.*, vol. 264, p. 125736, Mar. 2025, doi: 10.1016/j.eswa.2024.125736.

- [9]. M. N. Ali, E.-S. A. El-Dahshan, and A. H. Yahia, "Denoising of heart sound signals using discrete wavelet transform," *Circuits, Syst., Signal Process.*, vol. 36, no. 11, pp. 4482–4497, Nov. 2017, doi: 10.1007/s00034-017-0524-7.
- [10]. Macia-Sanahuja and H. Lamela-Rivera, "Wavelet analysis of partial discharges acoustic waves obtained using an optical fibre interferometric sensor for transformer applications," in *Proc. 2003 IEEE Int. Symp. Ind. Electron. (Cat. No. 03TH8692)*, vol. 2, 2003, pp. 1071–1076, doi: 10.1109/ISIE.2003.1267972.
- [11]. Z. K. Peng and F. L. Chu, "Application of the wavelet transform in machine condition monitoring and fault diagnostics: A review with bibliography," *Mech. Syst. Signal Process.*, vol. 18, no. 2, pp. 199–221, Mar. 2004, doi: 10.1016/S0888-3270(03)00075-X.
- [12]. X. Yu, W. Liang, L. Zhang, H. Jin, and J. Qiu, "Dual-tree complex wavelet transform and SVD based acoustic noise reduction and its application in leak detection for natural gas pipeline," *Mech. Syst. Signal Process.*, vol. 72–73, pp. 266–285, May 2016, doi: 10.1016/j.ymssp.2015.10.034.
- [13]. G. Wang, Y. Wang, Y. Min, and W. Lei, "Blind source separation of transformer acoustic signal based on sparse component analysis," *Energies*, vol. 15, no. 16, p. 6017, Aug. 2022, doi: 10.3390/en15166017.
- [14]. Y. Dong, H. Jiang, M. Mu, and X. Wang, "A trustworthy lightweight multi-expert wavelet transformer for rotating machinery fault diagnosis," *Mech. Syst. Signal Process.*, vol. 235, p. 112945, Jul. 2025, doi: 10.1016/j.ymssp.2025.112945.
- [15]. X. Hou, Y. Qian, and Y. Che, "Denoising method for transformer acoustic signal based on CEEMDAN-SVD-wavelet threshold," *J. Phys.: Conf. Ser.*, vol. 3087, no. 1, p. 012031, Aug. 2025, doi: 10.1088/1742-6596/3087/1/012031.
- [16]. T. Choudhary, V. Goyal, and A. Bansal, "WTASR: Wavelet transformer for automatic speech recognition of Indian languages," *Big Data Min. Anal.*, vol. 6, no. 1, pp. 85–91, Mar. 2023, doi: 10.26599/BDMA.2022.9020017.

Publisher's note: The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. The content is solely the responsibility of the authors and does not necessarily reflect the views of the publisher.

ISSN (Online): 3105-9082