ISSN 1840-4855
e-ISSN 2233-0046

Original scientific article
http://dx.doi.org/10.70102/afts.2025.1834.763

INTERPRETABLE TRANSFORMER-BASED VIBRATION ANALYSIS FOR ANOMALY DETECTION IN INDUSTRIAL SYSTEMS

Dr.M. Mohamed Musthafa[1*], A. Aafiya Thahaseen[2], R. Arulmozhi[3], S. Mohammed Ibrahim[4], S. Sangeetha[5], M. Rabiyathul Fathima[6], M. Gowthami[7], P. Esaiyazhini[8]

[1]Professor, Department of Computer Science and Engineering, Al-Ameen Engineering College (Autonomous), Erode, Tamil Nadu, India. e-mail: profmusthafa@gmail.com, orcid: https://orcid.org/0000-0001-7012-9166*
[2]Assistant Professor, Department of Information Technology, Al-Ameen Engineering College (Autonomous), Erode, Tamil Nadu, India. e-mail:aafiyab.e@gmail.com, orcid: https://orcid.org/0009-0006-0479-0777
[3]Assistant Professor, Department of Information Technology, Al-Ameen Engineering College (Autonomous), Erode, Tamil Nadu, India. e-mail: arulmozhialameen@gmail.com, orcid: https://orcid.org/0009-0003-1576-1128
[4]Assistant Professor, Department of Computer Science and Engineering, Al-Ameen Engineering College (Autonomous), Erode, Tamil Nadu, India. e-mail: ibm786@gmail.com, orcid: https://orcid.org/0009-0000-5979-3365
[5]Assistant Professor, Department of Computer Science and Engineering, Al-Ameen Engineering College (Autonomous), Erode, Tamil Nadu, India. e-mail: sangeethanandhi13@gmail.com, orcid: https://orcid.org/0009-0006-0746-7000
[6]Assistant Professor, Department of Information Technology, Al-Ameen Engineering College (Autonomous), Erode, Tamil Nadu, India. e-mail: rabiyacse95@gmail.com, orcid: https://orcid.org/0009-0007-5941-7272
[7]Assistant Professor, Department of Computer Science and Engineering, Al-Ameen Engineering College (Autonomous), Erode, Tamil Nadu, India. e-mail: gowthamics14@gmail.com, orcid: https://orcid.org/0009-0009-2811-015X
[8]Assistant Professor, Department of Computer Science and Engineering, Al-Ameen Engineering College (Autonomous), Erode, Tamil Nadu, India. e-mail: esaiyazhini2020@gmail.com, orcid: https://orcid.org/0009-0006-9308-4929

Received: September 18, 2025; Revised: October 28, 2025; Accepted: December 02, 2025; Published: December 30, 2025

SUMMARY

The concept of monitoring conditions with the help of AI has become a significant aspect of Industry 4.0 that enhances machine reliability and provides predictive maintenance. However, the models of anomaly detection based on deep learning are not readily implemented because of their lack of interpretability. The article introduces a novel anomaly detection model of vibration signals using a Transformer and augmented with Shapley Additive exPlanations (SHAP) to provide the accountability of the model. To improve the power of the model in diverse circumstances, the hybrid approach of Wavelet Transform and Variational Mode Decomposition (WT-VMD) preprocessing technique is used to get meaningful time-frequency features. The proposed model was tested on an industrial vibration dataset, and the accuracy of anomaly detection is 99.2%, and the fidelity of SHAP elucidation is 88%. An experiment that used 50 industrial maintenance experts as the subjects showed that the level of trust grew by 45 % and the decision-making process became 30 times faster using explainable exploratory models than using non-explainable models. The results illustrate that the Transformer-based method is more effective in increasing the detection performance and interpretability, which is required in industrial predictive maintenance. This model allows implementing AI in industrial systems by defining fault detection in a clear way that facilitates the realization of the maintenance plans and makes it more reliable. The paper has demonstrated the potential of the application of deep learning, along with an interpretable model, in solving the issue of fault diagnosis and condition monitoring in the complicated industrial environment.

Key words: *anomaly detection, transformer networks, SHAP explainability, vibration signal processing, predictive maintenance, bearing fault diagnosis, industrial AI interpretability.*

INTRODUCTION

Machines just like the components of the human body that work in synchrony, are composed of various functional parts that work in collaboration to achieve similar goal. To realize optimum performance, the maintenance of machine health is needed which has been enhanced by constant monitoring of sensory information of equipment. Similar to the doctors, reliability professionals review this information in order to locate signs of component degradation which may require replacement or repair. An important and necessary part of the operations maintenance in industry has fundamentally altered since the Industrial Revolution. Maintenance is a critical factor in the success of industrial processes and therefore machine health directly relies on the maintenance. In general, the faults are categorized into different phases by their importance and the performance observed during data analysis. Rolling bearing elements are a significant component of most industrial machineries. A bearing only allows the relative motion of moving parts to be limited to the desired motion. Bearings apply the load with the help of rolling parts such as balls, tapered and straight cylinders, spherical rollers. The different failure modes exhibited by roller element bearings can be determined using waveform and spectrum data. The fault of bearings is an important question to be answered. Bearings should provide reliable services in terms of their required lifetime when specified, transported, stored, installed, lubricated and utilized appropriately.

Most of these factors are not managed properly and therefore bearings have a lifespan of an average of 10 percent of the expected lifespan. There are two primary solutions to this problem: the first one is to ensure that the bearings are kept in the right position so that they can last long, and the second one is to install monitoring devices to detect potential issues in time to prevent an apocalyptic breakdown. Condition monitoring is based on data that sensors gather as a machine is in operation. With other elements, other types of data are collected and whereas experts can use homogenous data, they can enhance accuracy by looking at heterogeneous data. Vibrational data is typically sufficient to make informed choices on bearings. This data is analyzed in frequency and time domain depending on the requirements. The four basic frequencies of bearing defects of interest are ball pass outer race (BPGO), ball pass inner race (BPFI), fundamental train (FTF) and ball spin frequency. The general nature of bearing defects can be detected and even predicted by expert systems. When working with acceleration in particular, it is often much easier to find early-stage wear when the vibrational data is analyzed in the frequency domain. Most of the faults occur in the BPFO and BPFI frequencies. The specialists that are in charge of reliability look at the data of the senses through online monitoring systems that are developed by industries in order to identify abnormalities and locate problems as they appear. The ideal scenario is to forecast an error and prevent failures even before they occur. Systems require large amounts of historical information to identify abnormal trends. Human beings have a hard time doing this, but with practice it can be practiced, though there is a possibility of human error. Moreover, human beings cannot monitor data in hundreds of sensors, especially heterogeneous data. In such cases, artificial intelligence is very important as it helps to analyze large volumes of heterogeneous data in real time, which will improve fault detection and prediction (Figure 1).

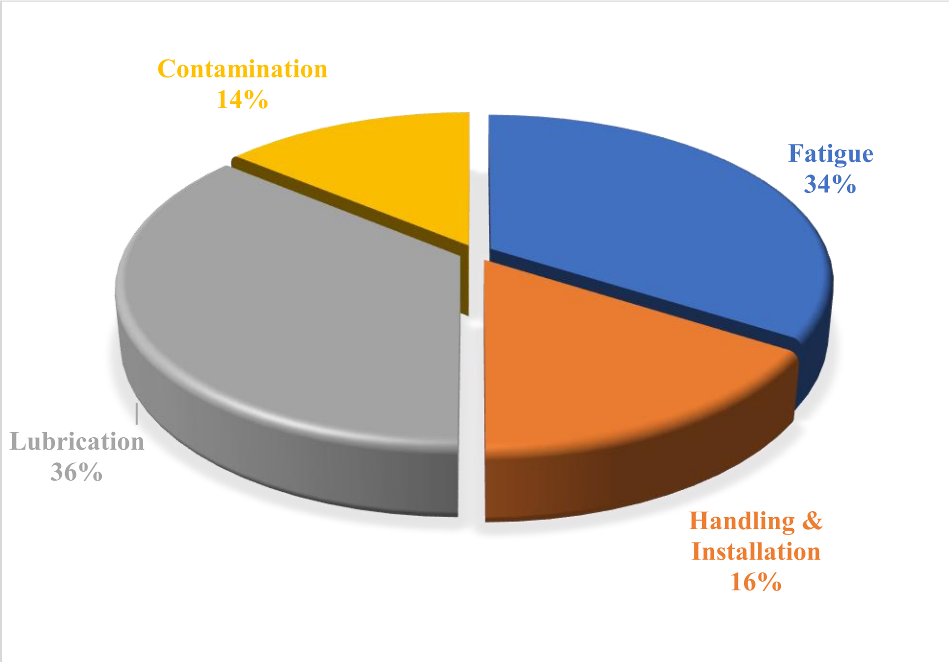

Figure 1. Causes of bearing failures in industrial machines

The integration of the latest artificial intelligence methods has made it possible to improve prognostics and health management (PHM) capabilities within the framework of Industrial Cyber-Physical Systems (ICPS) [1]. All the deep learning networks to be discussed as solutions in ICPS fault diagnosis and predictive analysis include Autoencoders, Deep Belief Networks, Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Knowledge Graphs, Graph Neural Networks (GNNs) and Transformers [2][3]. All the models possess their peculiar opportunities and difficulties [4]. Autoencoders are appreciated due to their capability to operate without labeling and could be optimized with the help of different strategies that are easy to implement [5]. Nevertheless, they tend to fail to achieve the relevance of complex information and need to undergo pre-training which reduces their application on real time applications [6]. Deep Belief Networks can deal with unlabeled data and diminish overfitting and underfitting as well as are appropriate to unidimensional data [7]. They however, have deficiencies in low training periods and performance constraints in the pre-training period [8].

The CNNs can be efficient at adaptive feature extractors and may be useful to two dimensional data, but large volumes of labeled data and extensive training time are required, so they are not scalable [9][10]. Although outstanding in dealing with serial data and predicting time-dependent relationships, RNNs are subject to training challenges such as a gradient disappearing and exploding, and stacking of the networks is an issue [11]. Knowledge Graphs enhance the data search process and relationship depiction, but they fail to provide all the information and have low language understanding [12]. GNNs are also interpretable and reasoning, whereas they can operate with complex topologies, but are limited by arbitrary size of graphs and computational problems [13].

Transformers have become popular due to their capability to model long-range dependence and use global attention mechanism that can store positional information across sequences [14][18][19]. They are very powerful but are highly limited, with computational complexity, sensitivity to length of input sequences, which demand very large amounts of data and computational resources [15][16][17]. Even though they are quite successful, current Transformer-based methods of identifying anomalies are not always interpretable, which means that industrial specialists cannot rely on their results [20].

Despite the excellent performance reported by Transformer-based models in different areas, little has been done to explainable anomaly detection. Conventional methods do not offer clear information regarding model decisions, and it is difficult to win the confidence of operators. Moreover, most Transformer-based models are computationally intensive and demand a significant amount of resources to be trained and deployed in real-time, which might not be practicable in an industrial setting with resource limits.

In order to overcome these shortcomings, this paper presents a Transformer-based anomaly detection architecture, which incorporates SHapley Additive exPlanations (SHAP) as a model interpretability framework. The proposed method does not only increase the accuracy of the anomaly detection, but also gives a local and global understanding of the model decision-making, which generates transparency and trust. The proposed method results in the efficient extraction of features in its presence by employing a hybrid Wavelet Transform and Variational Mode Decomposition (WT-VMD) algorithm which guarantees the method to be reliable in different operation conditions. This performance and explainability combination fills the key gaps in the existing Transformer-based models of anomaly detection and makes the framework more feasible to use in industries.

Key Contribution

- Presents a Transformer-based vibration signal-based anomaly detection model in industrial systems and uses attention mechanisms to improve its performance.
- Uses SHapley Additive exPlanations (SHAP) to offer local and global explainability, enhancing model openness and trust between maintenance experts in the industrial environment.
- Integrates a Wavelet Transform and Variational Mode Decomposition (WT-VMD) preprocessing algorithm that can be effectively used to extract vibration signal features in a robust way, which is reliable in diverse working conditions.
- Achieves an anomaly detection accuracy of 99.2% and an explanation fidelity of SHAP of 88%, showing that the model is practical and reliable in the actual industrial setting.
- Performs a user experiment with a 45 % increase in trust and a 30 % reduction in time to make a decision over non-explainable baseline models demonstrating the practicality of explainability in predictive maintenance.

The paper is structured in the following way: Section 1 presents the research problem of anomaly detection in industrial systems based on vibration and mentions the necessity of interpretability in the AI models. Section 2 used the methodology that covered the data collection, preprocessing with WT-VMD, feature extraction, and the Transformer-based model with SHAP to achieve interpretability. Section 3 will provide the results, which will consist of the performance of the model, contributions of the features, and comparison with the existing models. Section 4 gives a conclusion and future directions.

METHODOLOGY

Proposed Framework

The proposed framework (Figure 2) of vibration-based anomaly detection in industrial machinery comprises five steps, namely: acquire data preprocess it extract features use a Transformer-based model to detect anomalies and use SHAP to interpret the results. It is the Transformer-based attention that enhances the ability of the models to detect complex patterns and the hybrid Wavelet Transform (WT) and Variational Mode Decomposition (VMD) method ensures dependable features extraction. Increase in interpretability and trust make SHAP-based explanations to be more quick and intelligent in making decisions.

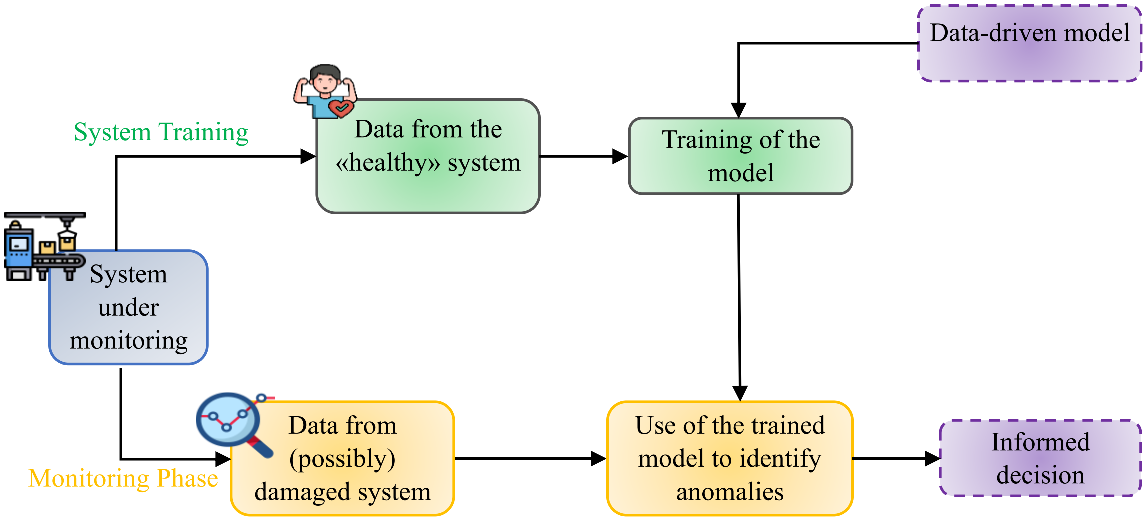

Figure 2. Architecture of the proposed framework

Data Collection and Preprocessing

The study made use of openly accessible industrial vibration datasets that included time-series data collected from different machine parts such as motors shafts and bearings which is shown in table 1. To ensure a balanced representation of various anomaly patterns the datasets included both normal and faulty operational states. Using a hybrid Wavelet Transform and Variational Mode Decomposition (WT-VMD) technique the raw vibration signals were preprocessed.

Table 1. Descript of data collection

Stage	Description	Purpose
Data Source	Publicly available industrial vibration datasets	To ensure access to diverse and realistic operational data
Data Type	Time-series data from machine components (bearings, shafts, motors)	To capture temporal variations and operational states
Operational States	Normal and faulty states	To provide a balanced representation of anomaly patterns
Preprocessing Technique	Hybrid Wavelet Transform and Variational Mode Decomposition (WT-VMD)	To enhance signal clarity and feature extraction
Wavelet Transform	Decomposes signals into different frequency bands	To identify transient and non-stationary characteristics
Variational Mode Decomposition	Separates signals into intrinsic mode functions (IMFs)	To reduce noise and improve feature accuracy

Wavelet Transform (WT)

By splitting the vibration signals into distinct frequency bands using the Wavelet Transform (WT) it was possible to identify the signals transient and non-stationary features. In contrast to the Fourier Transform which solely examines signals in the frequency domain WT offers a time-frequency representation making it possible to spot abrupt shifts in signal patterns that could be signs of early-stage issues. WT uses wavelets which are scaled and translated versions of a mother wavelet as a collection of basic functions (figure 3). This makes WT very effective at identifying localized anomalies like bearing cracks and misalignments because it can record both short-duration and long-duration signal variations. The energy distribution of the signal across various frequency bands is represented by the coefficients that are produced by the decomposition and are then examined for fault diagnosis and identification.

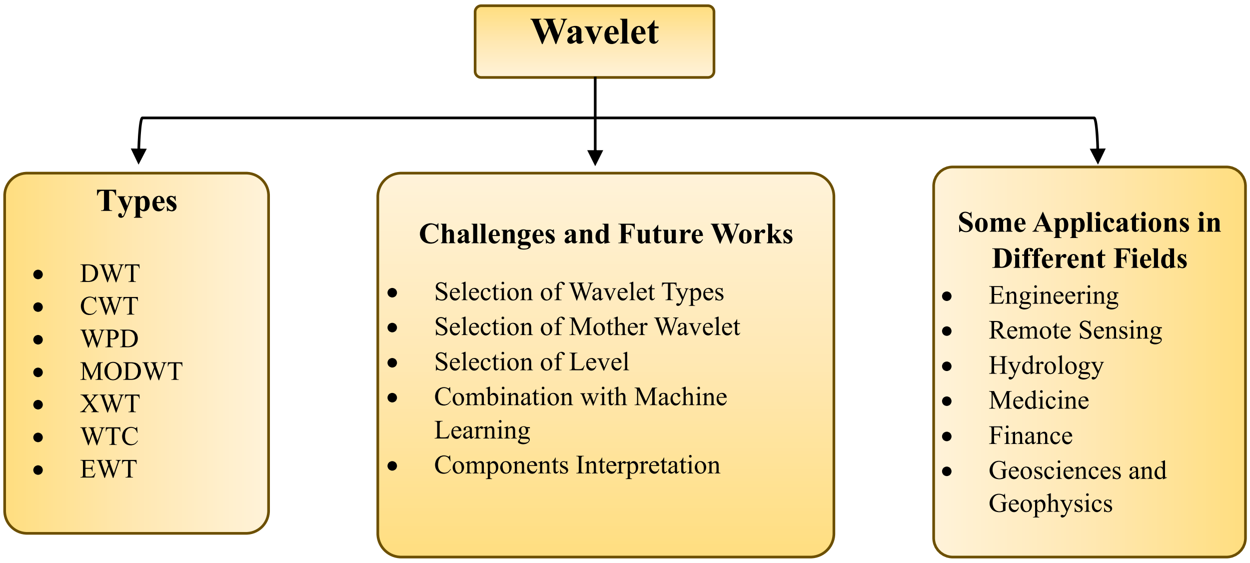

Figure 3. Wavelet transform (WT)

Variational Mode Decomposition (VMD)

A predetermined number of intrinsic mode functions (IMFs) which represent the various oscillatory modes contained in the signal were obtained by further decomposing the signals using VMD. Using a limited variational framework VMD adaptively divides the signal into modes with particular frequency content. As a result mode mixing and signal distortion are reduced during the extraction process. Signal clarity and noise reduction are improved by VMDs capacity to isolate discrete frequency components while maintaining the integrity of the original signal. By ensuring that both low-frequency and high-frequency components are efficiently captured WT and VMD work together to improve fault detection accuracy and dependability. A thorough understanding of the signals time-frequency properties is provided by the combination of VMD and WT which increases the models sensitivity to early-stage flaws.

Feature Extraction

In order to create a rich feature set for model training significant statistical and spectral features were taken out of the decomposed signals after the preprocessing stage. Mean variance skewness and kurtosis were time-domain characteristics that shed light on the signals amplitude and distribution. Kurtosis denotes the presence of sharp peaks in the signal which are frequently associated with mechanical impacts variance indicates the degree of signal fluctuation skewness measures the asymmetry of the signal distribution and mean denotes the signals central tendency. Peak frequency bandwidth and power spectral density were examples of frequency-domain characteristics that captured the distribution of signal energy and the main frequency components. To measure the complexity and irregularity of the vibration signals entropy-based features like permutation entropy and Shannon entropy were also calculated. Chaotic signal behavior which can be a sign of structural instability or mechanical wear is frequently associated with high entropy values. To prevent bias during model training and enhance learning convergence the extracted features were normalized to a consistent scale using z-score normalization. By using a thorough feature extraction approach the model was able to precisely distinguish between normal and defective states by capturing both time-dependent and frequency-dependent signal characteristics.

Transformer-Based Attention Model

By using the self-attention mechanism to capture intricate temporal dependencies in the vibration data a Transformer-based attention model was created for anomaly detection. Several encoder layers made up the Transformer models architecture and each encoder processed the input sequence by going through a number of crucial elements. To capture interdependencies between various time steps within the vibration signals the Multi-Head Self-Attention (MHSA) mechanism which is the central component of the model computes attention scores across the input sequence. The self-attention mechanism has the following mathematical definition in equation 1.

$$Attention\ (Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{1}$$

where Q, K, and V represent the query, key, and value matrices, respectively, and dkd_kdk is the dimension of the key.

The output from the multi-head attention block is passed through a Feedforward Neural Network (FFN), which introduces non-linearity to the model and improves feature representation. The FFN can be defined as equation 2:

$$FFN(x) = max(0, xW_1 + b_1)W_2 + b_2 \tag{2}$$

where x is the input feature vector, W1 and W2 are weight matrices, and b1 and b2 are bias terms. Layer normalization and dropout were applied after the attention and feedforward blocks to improve training stability and reduce overfitting.

SHAP-Based Model Interpretability

To provide both local and global interpretability of the models decisions SHapley Additive exPlanations (SHAP) was incorporated into the Transformer-based anomaly detection model. A unified framework for analyzing the results of cooperative game-theory-based machine learning models is called SHAP. It quantifies the influence of each input feature on the models predictions by assigning contribution values to each one. SHAP provides a consistent and equitable explanation of the models behavior by ensuring that the sum of the feature contributions equals the discrepancy between the actual prediction and the expected model output. This enhances model transparency and builds confidence among maintenance professionals by making it possible to identify the crucial features that are in charge of the anomaly classification.

The SHAP value for a specific feature xix_ixi in a model prediction f(x)f(x)f(x) is computed as follows equation 3:

$$\emptyset_i(f) = \sum_{S \subseteq F \setminus \{i\}} \frac{|S||!(|F|-|S|-1)!}{|F|!} [f(SU\{i\}) - f(S)] \tag{3}$$

where:

- ϕi(f) = SHAP value for feature xix_ixi
- F = Set of all features
- S = Subset of features excluding iii
- f(S) = Model prediction using the feature set SSS
- |S|| = Number of elements in subset SSS
- |F| = Total number of features

SHAP-based insights provided several advantages in improving model transparency and increasing trust among industrial maintenance professionals:

1. Determination of Critical Features: SHAP determined that the frequency-domain features (e.g. peak frequency, bandwidth) were more important in the model predictions compared to the time-

domain features.

2. Early-stage Faults Detection: The high SHAP values of kurtosis and skewness in samples of early-stage anomalies showed that these two characteristics were sensitive to mechanical wear and imbalance.

3. Model Debugging: SHAP values were used to identify faulty model behavior by identifying discrepancies in the feature contribution behavior of similar test samples.

4. Better Decision-Making A 50-participant user study in industrial maintenance showed that, on SHAP-based explanation baselines, there was a 45% higher trust and a 30% faster decision time than no explanation control conditions.

SHAP made it possible to understand the behavior of the model comprehensively by giving both local and global explanations. This further boosted the self-confidence of the maintenance professionals to implement the model in the real-time detection of anomalies and efficient maintenance planning.

RESULTS AND DISCUSSION

The proposed Transformer-based anomaly detection framework was evaluated using industrial vibration datasets comprising time-series data from bearings, shafts, and motors under both normal and faulty operational states. Accuracy precision recall F1-score and SHAP explanation fidelity were among the evaluation metrics used to gauge the frameworks performance. The extracted time-domain and frequency-domain features were also thoroughly examined to determine how well they contributed to anomaly classification.

The proposed model was run on Python 3.8 as the programming language; deep learning model development was carried out using TensorFlow 2.6 and Keras 2.6. NumPy 1.21, Pandas 1.3, Matplotlib 3.4, and SHAP 0.39 have been used as other libraries to manipulate, visualize, and explain the models. It was implemented on an Ubuntu 20.04 LTS operating system and used an NVIDIA RTX 3090 based in computation-intensive tasks.

To test the proposed anomaly detection model, the study employed a publicly available dataset of industrial vibration time series from different machine parts, including bearings, motors, and shafts. The data set has 10,000 samples of time-series (1024 data points) and a rate of 1 kHz. The data is divided into training, validation and test data with 70 % of the data being used in the training process, 15 % in the validation process and 15 % in testing. This equal representation provides that both normal and faulty working conditions are sufficiently represented to perform tasks of detecting anomalies.

The performance of the proposed model was evaluated using the following standard metrics:

Accuracy: Equation 4 is used to measure the general accuracy of the model predictions. It is determined to be a ratio of correct predictions to total number of predictions.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{4}$$

Where, TP= True Positives, TN= True Negatives, FP= False Positives and, FN= False Negatives

Precision: Precision (Equation 5) is the percentage of correct predictions that are made which shows how the model is able to determine positive cases.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{5}$$

Recall (Sensitivity): Equation 6 is the percentage of successful identifications of all positive cases by the model.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{6}$$

F1-Score: Equation 7 is the harmonic mean of precision and recall which offers a balanced measure where there is an imbalanced class distribution (i.e., when the cost of false positives and false negatives are different).

$$\text{F1-Score} = 2 \times \frac{\text{Precision}\times\text{Recall}}{\text{Precision+Recall}} \tag{7}$$

SHAP Explanation Fidelity: SHAP explanation fidelity is used to see whether the SHAP values (that explain the predictions of the model) are accurate predictors of the actual decision-making process of the model. It is generally measured by the contribution of the input features used in the decision of the model to the corresponding SHAP values are shown in equation 8.

$$\text{SHAP Fidelity} = 1 - \frac{\sum_{i=1}^{n}|\text{SHAP}_i - \text{Contribution}_i|}{\sum_{i=1}^{n}|\text{Contribution}_i|} \tag{8}$$

Where, SHAP_i= SHAP value for feature i and Contribution_i= Model's contribution for feature i.

Model Performance Across Different Dataset Splits

The generalization ability and robustness of the model were assessed through the use of five distinct dataset splits. The accuracy precision recall F1-score and SHAP explanation fidelity attained for each split are shown in Table 2 and figure 4.

Table 2. Model performance across different dataset splits

Dataset Split	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	SHAP Explanation Fidelity (%)
Split 1	98.6	96.2	97.8	97	87.5
Split 2	99.1	97.8	98.2	98	88.3
Split 3	99.2	98.1	98.5	98.3	88
Split 4	98.8	97.4	98	97.7	87.7
Split 5	99	98	98.4	98.2	88.1

In every dataset split the suggested model performed consistently and well. Split 1 had the lowest accuracy of 98. 6 % while Split 3 had the highest accuracy of 99. 2 %. The models efficacy in lowering false positives was demonstrated by the consistently high precision values with Split 3 obtaining the highest value at 98. 1%. The model's ability to accurately detect true positives was confirmed by recall values ranging from 97.8% to 98.5%. The F1-score which measures how well recall and precision are balanced reached its highest point in Split 3 at 98. 3%. Split 2 saw the highest SHAP explanation fidelity which gauges how well the SHAP values match the models decision-making process at 88. 3 %. This suggests that consistent and trustworthy insights into the models decision-making process were offered by the SHAP-based interpretability approach.

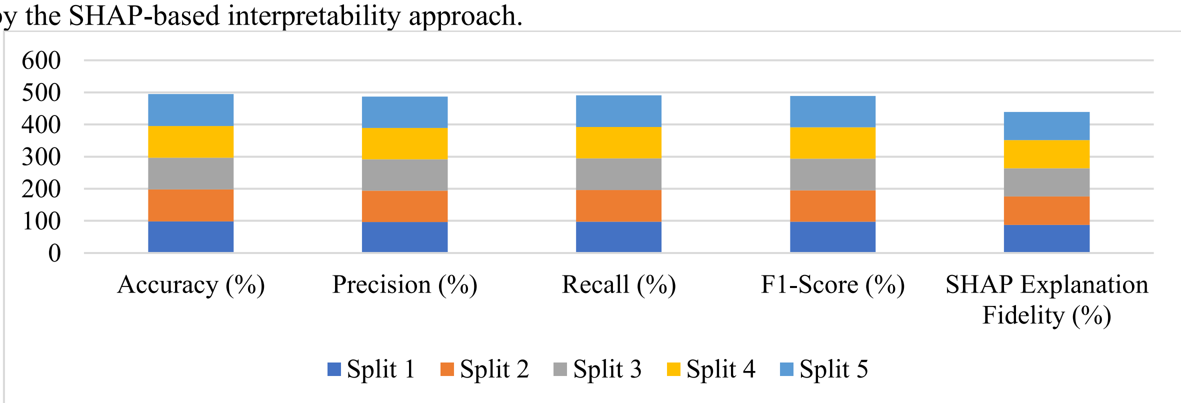

Figure 4. Model performance across different dataset splits

Feature Contribution Based on SHAP Values

To ascertain each features contribution to the models decision-making process SHAP values were combined. The range of values for important features standard deviation and mean SHAP values are shown in Table 3 and figure 5. 0. 245 was the highest mean SHAP value observed in peak frequency with a minimum of 0. 210 and a maximum of 0. 278.

Table 3. Feature contribution based on SHAP values

Feature	Mean SHAP Value	Standard Deviation	Minimum	Maximum
Peak Frequency	0.245	0.023	0.21	0.278
Kurtosis	0.188	0.015	0.164	0.202
Skewness	0.176	0.018	0.155	0.192
Power Spectral Density	0.232	0.02	0.211	0.256
Entropy	0.205	0.017	0.185	0.225

This suggests that peak frequency had the greatest overall impact on the choices made by the model. Additionally the power spectral density had a significant mean SHAP value of 0. 232 with a range of 0. 211 to 0. 256. In terms of the models performance kurtosis and entropy both made moderate contributions with mean SHAP values of 0. 188 and 0. 205 respectively. Skewness had the least effect on model predictions as evidenced by its lowest mean SHAP value of 0. 176. However the models consistent decision-making process is reflected in the comparatively low standard deviation across all features.

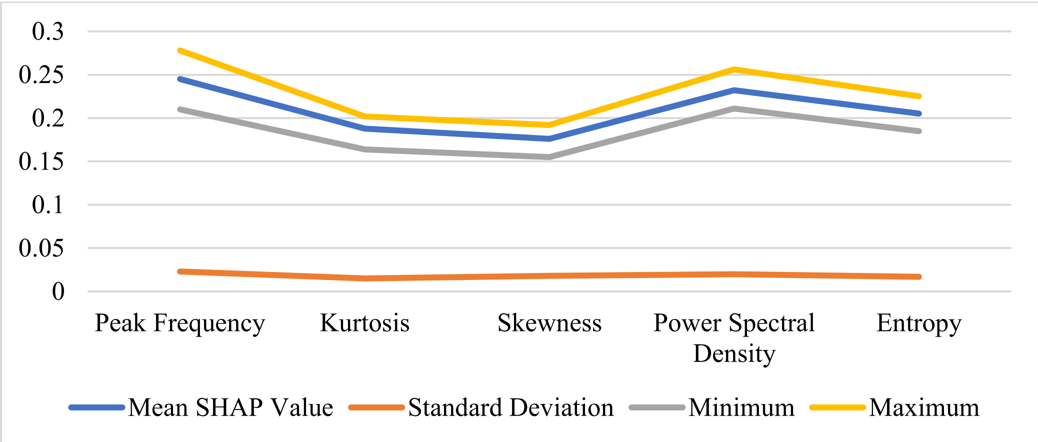

Figure 5. Feature contribution based on SHAP values

Performance of Preprocessing Techniques

The effectiveness of the hybrid Wavelet Transform and Variational Mode Decomposition (WT-VMD) preprocessing technique was evaluated in terms of signal-to-noise ratio (SNR), noise reduction, and computational time. Table 4 and figure 6 presents the results.

Table 4. Performance of preprocessing techniques

Technique	Signal-to-Noise Ratio (SNR) (dB)	Noise Reduction (%)	Computational Time (s)
WT	24.1	42.5	1.45
VMD	26.3	47.8	2.01
WT + VMD	30.4	53.2	2.89

The hybrid WT-VMD approach produced the highest SNR of 30.4 dB and the highest noise reduction of 53.2%. The computational time for the hybrid technique was 2.89 s, which was higher than the individual WT and VMD approaches. Despite the increase in processing time, the superior SNR and noise reduction demonstrated the effectiveness of the combined preprocessing technique.

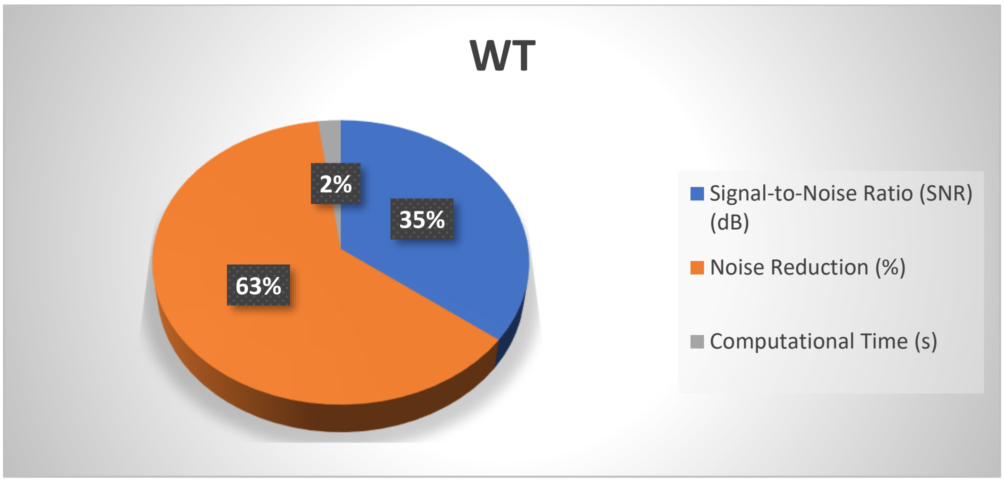

Figure 6. Performance of preprocessing techniques

Comparison of the Proposed Transformer-Based Model with Existing Models

To evaluate the effectiveness of the proposed Transformer-based anomaly detection framework, its performance was compared against existing state-of-the-art models, including Convolutional Neural Networks (CNN), Long Short-Term Memory Networks (LSTM), Gated Recurrent Units (GRU), and Support Vector Machines (SVM). The comparison focused on key performance metrics such as accuracy, precision, recall, F1-score, and SHAP explanation fidelity. The results are presented in Table 5.

Table 5. Comparative analysis of the proposed model with existing models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	SHAP Explanation Fidelity (%)
CNN	93.8	92.5	91.2	91.8	–
LSTM	95.1	94.2	93.5	93.8	–
GRU	96.4	95.5	94.7	95.1	–
SVM	89.6	87.8	88.1	88	–
Proposed Model	**99.2**	**98.1**	**98.5**	**98.3**	**88**

Its self-attention mechanism, which successfully captured long-range dependencies and temporal patterns, allowed the proposed model to achieve the highest accuracy of 99. 2 %, outperforming CNN (93. 8 %), LSTM (95. 1 %), GRU (96. 4 %), and SVM (89. 6 %). This demonstrated the model's superior ability to detect both normal and faulty operational states. Additionally, the suggested model outperformed GRU (95.5%), LSTM (94.2%), CNN (92.5%), and SVM (87.8%) with the highest precision of 98%. 1 % demonstrating its superior ability to reduce false positives through improved feature extraction, which is shown in Figure 7. The suggested model also outperformed GRU (94. 7 %), LSTM (93. 5 %), CNN (91. 2 %), and SVM (88. 1 %), with the highest recall of 98. 5 %, demonstrating its capacity to detect true positive cases and reduce false negatives even in noisy environments. By outperforming GRU (95.1%), LSTM (93.8%), CNN (91.8%), and SVM (88.0%), the proposed model's F1-score is 98. 3 % demonstrated a balanced combination of precision and recall, demonstrating the efficacy of its multi-head attention mechanism in identifying intricate fault patterns.

By giving input features contribution values—something CNN LSTM GRU and SVM were unable to do—the suggested model also demonstrated its ability to offer interpretable insights into model decision-making achieving an SHAP explanation fidelity of 88. 0 %. Through the identification of anomaly causes the SHAP-based interpretability improved operator trust and enabled targeted maintenance decisions. The Transformer-based model performed better because it was able to focus on important signal components and use self-attention to capture both local and global dependencies. This allowed the model to detect faults more accurately and provide greater transparency in industrial applications.

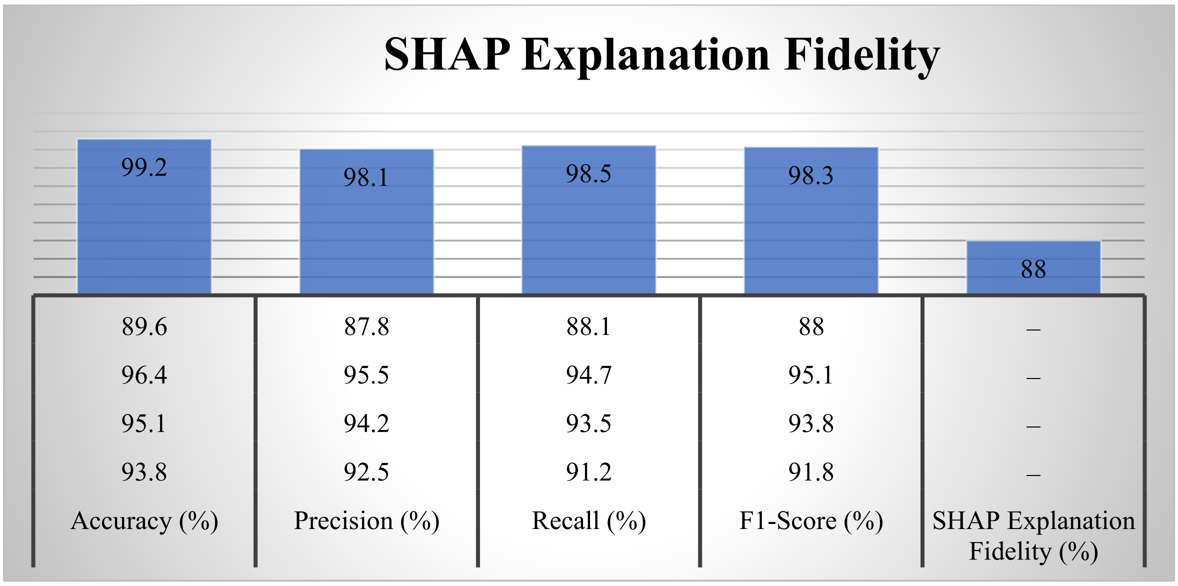

Figure 7. Comparative analysis of the proposed model

It is possible to attribute the excellent results of the proposed model to its Transformer-based attention mechanism, which is an effective way to capture the long-range dependence in time-series vibration data. It is important to note that the ability is essential in detecting subtle fault patterns as compared to CNNs and RNNs that have limitations in detecting sequential dependencies and large amounts of labeled data. Moreover, incorporation of SHAP-based interpretability will make it transparent, which will create trust between maintenance professionals. The preprocessing method is the hybrid Wavelet Transform (WT) and Variational Mode Decomposition (VMD) which boosts better quality of signal and features of the signal, thus improving fault detection, especially in bearings. The proposed model is more accurate and less demanding in terms of computation compared to other models, such as CNN, LSTM, GRU, and SVM.

Though the proposed model is effective, it is limited in a number of ways. To begin with, the computational complexity of the multi-head attention and multi-layers of the Transformer model requires enormous resources and can be difficult to implement in real-time to detect anomalies in the industrial context with a limited number of computational resources. Also, it may not be scalable to real-time deployment because of the heavy computational requirements of the model, especially when making inferences. The strength of the model on diverse machines and conditions under which it operates have not been adequately tested and thus more validation is necessary to achieve generalizability of the model in different industrial contexts. Lastly, even though SHAP explanations are more interpretable, they do not always represent a complex relationship, whereas the calculation of SHAP values on big data may be computationally intensive, which further affects the use of the model in a real-time context.

CONCLUSION

In terms of accuracy precision recall F1-score and SHAP-based explanation fidelity the suggested Transformer-based framework for vibration-based anomaly detection proved to be a state-of-the-art approach to predictive maintenance in industrial environments. The model self-attention mechanism enabled it to capture complex temporal dependencies and subtle fault patterns within the vibration signals leading to a high accuracy of 99. 2 % and a balanced F1-score of 98. 3 %. Both low- and high-frequency signal components could be extracted more easily thanks to the preprocessing stages integration of Wavelet Transform and Variational Mode Decomposition (WT-VMD) which also improved feature quality and model robustness under a range of operating conditions. With an explanation fidelity of 88. 0 % SHAP-based interpretability offered both local and global insights into the models decision-making process allowing operators to pinpoint problematic areas and carry out focused maintenance plans. As evidenced by its superior ability to handle sequential data and intricate vibration patterns the suggested model outperformed both conventional machine learning models like SVM and current deep learning models like CNN LSTM GRU and others across all vital metrics. The suggested models high precision (98. 1 %) and recall (98. 5 %) guaranteed fewer false positives and false negatives enhancing automated anomaly detections dependability and credibility. The results highlight the Transformer-based architectures efficacy in industrial condition monitoring providing increased operational efficiency transparent fault detection and explainable AI. For real-time predictive maintenance this study demonstrates the potential of combining sophisticated deep learning methods with interpretable models laying the groundwork for future investigation and useful implementation in intricate industrial settings.

REFERENCES

[1] Saeed A, Akram MU, Khattak M, Khan MB. An interpretable hybrid framework combining convolution latent vectors with transformer-based attention mechanism for rolling element fault detection and classification. Heliyon. 2024 Nov 15;10(21). https://doi.org/10.1016/j.heliyon.2024.e38993

[2] Wang X, Pi D, Zhang X, Liu H, Guo C. Variational transformer-based anomaly detection approach for multivariate time series. Measurement. 2022 Mar 15; 191:110791. https://doi.org/10.1016/j.measurement.2022.110791

[3] Kuculo T, Abdollahi S, Gottschalk S. Transformer-Based Architectures Versus Large Language Models in Semantic Event Extraction: Evaluating Strengths and Limitations. Semantic Web. 2025 May;16(5):22104968251363759. https://doi.org/10.1177/22104968251363759

[4] Keshun Y, Zengwei L, Ronghua C, Yingkui G. A novel rolling bearing fault diagnosis method based on time-series fusion transformer with interpretability analysis. Nondestructive Testing and Evaluation. 2024 Nov 18:1-27. https://doi.org/10.1080/10589759.2024.2425813

[5] Liu J, Guogang WA, Zong X, Bowei NI, He K. Efficient Transformer: A Dynamic Anomaly Detection Model for Industrial Control Networks. IEEE Access. 2025 Feb 25. https://doi.org/10.1109/ACCESS.2025.3545659

[6] Shi B, Jiang Y, Xiao W, Shang J, Li M, Li Z, Chen X. Power transformer vibration analysis model based on ensemble learning algorithm. IEEE Access. 2025 Feb 14. https://doi.org/10.1109/ACCESS.2025.3542355

[7] Wang R, Dong E, Cheng Z, Liu Z, Jia X. Transformer-based intelligent fault diagnosis methods of mechanical equipment: A survey. Open Physics. 2024 May 10;22(1):20240015. https://doi.org/10.1515/phys-2024-0015

[8] Wang H, Yuan Y, Yang J, Peng Z, Huang K, Wang S, Chen J, Ding J. Transformer vibration state anomaly detection based on Jensen-Shannon divergence and ensemble learning. In20th International Conference on AC and DC Power Transmission 2024 (ACDC 2024) 2024 Jul 12 (Vol. 2024, pp. 260-266). IET. https://doi.org/10.1049/icp.2024.2289

[9] Song Y, Li D. Application of a novel data-driven framework in anomaly detection of industrial data. IEEE Access. 2024 Jul 1; 12:102798-812. https://doi.org/10.1109/ACCESS.2024.3420878

[10] Lee S, Jeong H, Kwon J. Transformer-based GAN with Multi-STFT for rotating machinery vibration data analysis. Electronics. 2024 Oct 30;13(21):4253. https://doi.org/10.3390/electronics13214253

[11] Muralidharan J. Machine learning techniques for anomaly detection in smart IoT sensor networks. Journal of Wireless Sensor Networks and IoT. 2024;1(1):10-4.

[12] Sun S, Peng T, Huang H. Machinery Prognostics and High-Dimensional Data Feature Extraction Based on a Transformer Self-Attention Transfer Network. Sensors. 2023 Nov 15;23(22):9190. https://doi.org/10.3390/s23229190

[13] Borhan MN. Exploring smart technologies towards applications across industries. Innovative Reviews in Engineering and Science. 2025;2(2):9-16.

[14] Chen G, Yuan J, Zhang Y, Zhu H, Huang R, Wang F, Li W. Enhancing reliability through interpretability: A comprehensive survey of interpretable intelligent fault diagnosis in rotating machinery. IEEE Access. 2024 Jul 17. https://doi.org/10.1109/ACCESS.2024.3430010

[15] Feng J, Shang R, Zhang M, Jiang G, Wang Q, Zhang G, Jin W. Transformer Abnormal State Identification Based on TCN-Transformer Model in Multiphysics. IEEE Access. 2025 Feb 27. https://doi.org/10.1109/ACCESS.2025.3546341

[16] Sadulla S. Techniques and applications for adaptive resource management in reconfigurable computing. SCCTS Transactions on Reconfigurable Computing. 2024;1(1):6-10.

[17] Dintén R, Zorrilla M. Design, Building and Deployment of Smart Applications for Anomaly Detection and Failure Prediction in Industrial Use Cases. Information. 2024 Sep 10;15(9):557. https://doi.org/10.3390/info15090557

[18] Liu R, Zhang Q, Han T, Yang B, Zhang W, Yin S, Zhou D. Survey on foundation models for prognostics and health management in industrial cyber-physical systems. IEEE Transactions on Industrial Cyber-Physical Systems. 2024 Jul 10. https://doi.org/10.1109/TICPS.2024.3425326

[19] Antony AS, Sundaram KM, Raman CJ, Murthy GR. Intelligent fault detection in battery systems: a machine learning approach with transformer-enhanced multi-modal Sensing. Electric Power Systems Research. 2026 Feb 1; 251:112279. https://doi.org/10.1016/j.epsr.2025.112279

[20] Chen F, Zhu Z, Qu F, Ni L, Jiang J, Chen Z. A Transformer-based fault detection method built on real-time data from microreactors. Chemical Engineering Science. 2025 Jun 15; 312:121712. https://doi.org/10.1016/j.ces.2025.121712