

Fault Detection in Spur Gears Through Vibration Signal Analysis and Machine Learning Techniques

Muhammad Talha Asif^{1,a}, Shahzad Ahmad^{2,b*}, Asif Khan^{3,c}, Sohail Malik^{4,d}

¹Faculty of Mechanical Engineering, Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Topi-23640, Swabi, Khyber Pakhtunkhwa, Pakistan

²Faculty of Mechanical Engineering, Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Topi-23640, Swabi, Khyber Pakhtunkhwa, Pakistan

³Mechanical Engineering Department, Physical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Kingdom of Saudi Arabia

⁴Mechanical Manufacturing and Automation, Pak-Austria Fachhochschule Institute of Applied Sciences and Technology, Mang Haripur, Pakistan

^atalha.asif@giki.edu.pk, ^{b*}shahzad.ahmad@giki.edu.pk, ^ckhanuet11@gmail.com,
^dsohail.malik@fcm3.paf-iast.edu.pk

Keywords: Short-Time Fourier Transform, Scalograms, Machine Learning, Fault Diagnosis, Gears Condition Monitoring.

Abstract. This research uses a multi-domain technique to give a thorough analysis of mechanical gears health evaluation that includes time, frequency, and time-frequency signal analysis. The research seeks to discover patterns indicative of healthy, partially damaged, or fully damaged gear states using a variety of graphical representations, including time and frequency plots, the Short-Time Fourier Transform (STFT), and scalograms which are visual representations of the wavelet transform of a signal. Advanced machine learning models are used to improve diagnostic accuracy when manual identification of these trends becomes difficult. The goal is to achieve a validation accuracy greater than 70% a threshold selected based on prior studies indicating that this level ensures reliable fault detection for industrial applications while balancing computational constraints. The reliability and effectiveness of gear monitoring systems can be increased by integrating contemporary signal processing and machine learning approaches, as demonstrated by this research, which also advances the identification of gear faults. Among the conclusions are the outcomes of tests done to identify gear problems in which authors were able to train a model with more than 72% accuracy and able to propose Artificial Intelligence model for classification of faults in gears.

1. Introduction

Mechanical gears are a fundamental part in many industrial applications such as automotive, aerospace, and manufacturing systems [1]. Failure to this part can lead to costly downtime, operational inefficiencies, and potential safety hazards [2]. So, to avoid these unexpected failures, the condition monitoring of gears has become a critical area of research in which vibration signal analysis is widely recognized for its ability to detect anomalies in gear operation [3]. Traditional fault detection approaches, such as time-domain and frequency-domain analyses struggle to identify early-stage defects due to the complex nature of vibration signals [4]. While techniques like Fast Fourier Transform (FFT) and Short-Time Fourier Transform (STFT) provide frequency-based insights and may fail to capture non-stationary signal variations. So, as a result, machine learning (ML) techniques have emerged as a powerful alternative, enabling automated fault detection by recognizing patterns that are not easily distinguishable through manual inspection [5]. A time frequency analysis was applied in industrial bearing and a pattern was observed for fault with a specific behavior but lacks accuracy at specific instants [6]. Vibration signals and thermal images were used for CNN, a type of deep learning model, but the model failed at some instant when there are uneven surrounding interventions [7]. At different variable working conditions signal processing, machine learning techniques and deep learning models were implemented, but the model needs high computational

power and difficult to implement at industrial sites [8]. Combination of 1D and 2D CNN used in Power Quality Disturbance (PQD) classification but requires processing time for training of neurons [9]. Hybrid Genetic Feature Selection (HGFS) framework was developed but lacks computational efficiency and requires a lot of processing time [10]. Recent studies, such as [9] have explored the use of deep reinforcement learning for real-time gear fault diagnosis, achieving higher accuracy but still facing challenges in computational efficiency. A web application that takes machine sensor data as input to predict the possibility of downtime of the machine and it needs good accuracy of the model to predict the life of the gears [11]. Based on previous literature, the current trends have a limitation of computational storage as with the adopted implementation of deep learning and feature extraction it requires a much computational power in the form of Graphical Processing Unit [12]. Also, it will not be feasible for real-time monitoring. So based on these, the simplified approach needs to be adopted in which a machine learning-based fault diagnosis framework needs to implement that integrates signal processing and supervised learning for enhancing the accuracy of gear condition monitoring [13]. Different feature extraction and selection techniques need to be implemented to optimize the dataset and train a classification model to distinguish between different conditions in gear and predict their health through predictive maintenance [14].

Based on the previous work it is desired to do fault diagnosis of gears through a simplified approach which can also be implemented in Internet of Things. Therefore, in this research, a setup is fabricated in which meshing of gears performed and data is extracted through sensor and different signal processing techniques were implemented in it. Finally, supervised learning is performed through which AI model obtained and implemented for real-time diagnosis of gears.

2. Methodology

The overall methodology includes a data acquisition process for the meshing of gears section. Arduino Mega 2560 Rev3 acquired that data and then processed further in MATLAB for fault detection in gears as shown in Figure 1.

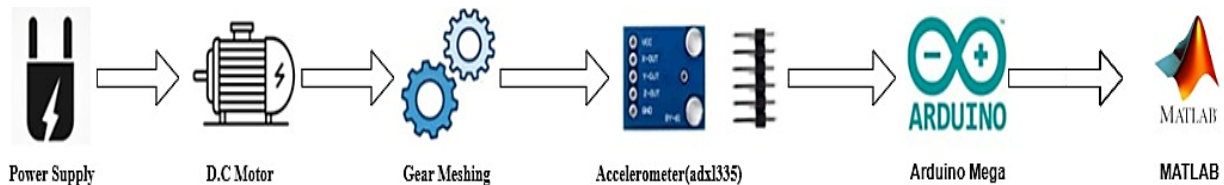


Fig. 1. Workflow of Experiment

2.1. Experimental Setup

The experimental setup used in this study is a gear-driven mechanical system with two shafts in which one shaft is attached to a DC motor directly while the other shaft is free at both ends. Different voltages were applied to the D.C motor and meshing of the gears were performed for different conditions of the spur gears. Initially three different classifications were implemented in the gears. Containing Healthy Gears in which the spur gears are in their healthy state, Partial Broken Teeth in which one of the teeth is broken and Partially Broken Gear in which a complete tooth is broken as shown in Figure 2.

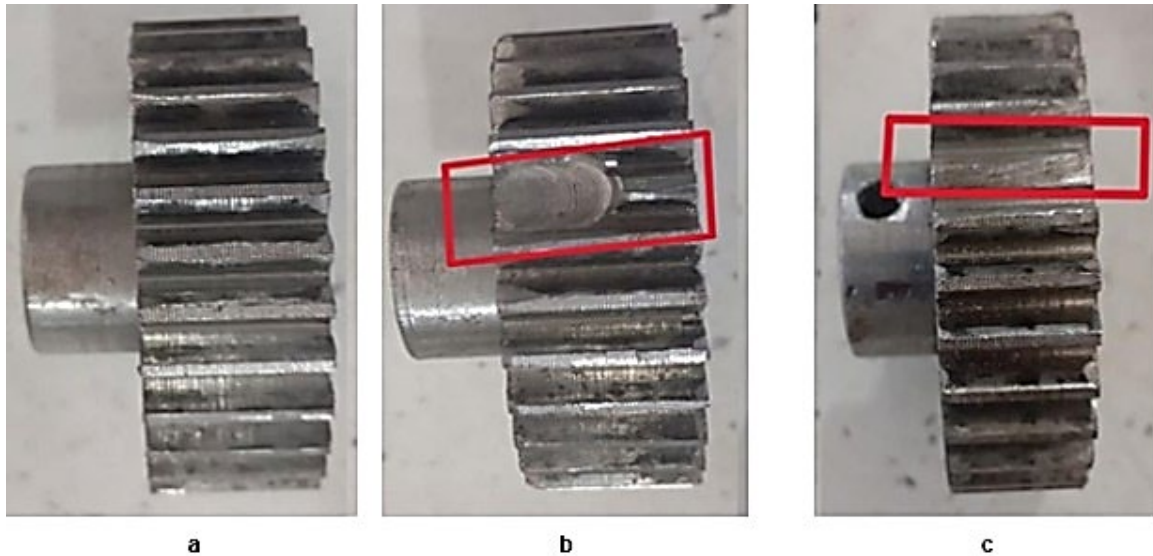


Fig. 2. a. Healthy(H) b. Partially Broken Gear (PBG) c. Partially Crack Teeth (PCT)

The experimental setup is such that gears are attached to the shafts aligned parallel to each other and rotation of one shaft causes the meshing of the two spur gears. The shafts are passed through bearing housing in which an accelerometer is mounted that is responsible for capturing vibration of gears. The data is captured on Arduino Mega 2560 Rev3 and further processed in SUMILINK. The Arduino Mega 2560 Rev3 was selected for its high analog input resolution (10-bit ADC), sufficient digital I/O pins (54), and compatibility with Simulink for real-time data acquisition, making it more suitable than alternatives like Arduino Uno or Raspberry Pi [15]. The experimental setup is shown in Figure 3.

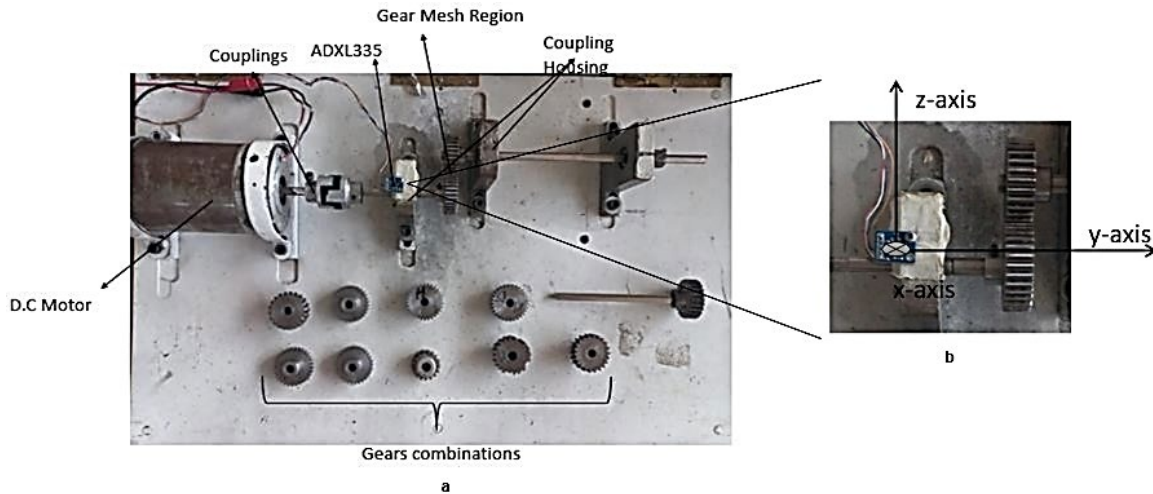


Fig. 3. a. Experimental Setup b. Accelerometer Axis

The accelerometer (ADXL335) is fixed to the bearing housing and records vibration on two axis that is x-axis and z-axis along the third axis where vibration is zero as it is along the shaft, so there will be minimum vibration in that axis. The data is acquired from only x-axis and z-axis to record maximum vibrations of the bearing housing generated through the gear meshing.

2.2. Data Acquisition

To gather and analyze data in real-time for this investigation, the Arduino Mega 2560 Rev3 microcontroller was integrated with the ADXL335 accelerometer and Simulink. The sampling time was adjusted to remove aliasing errors in it and for its Nyquist frequency criteria is implemented to set the sampling rate equal to twice the operating frequency [16]. The data is acquired at three distinct voltage levels that are at 10V, 15V, and 20V. These voltages chosen to simulate a range of operational

speeds typical in industrial gear systems, capturing diverse dynamic behaviors [17]. The Simulink model ran for 10 seconds for each voltage level, recording 1001 data points for the vibrations of the X-axis and Z-axis. The dataset was divided into 70% for training, 15% for validation, and 15% for testing, as this arrangement offered enough resolution to record the mechanical vibrations produced by the shafts and gears under various operating circumstances. Signal processing and fault categorization were then performed on the data [18].

2.3. Signal Processing

To examine and contrast the behavior of the gear system under the three operating conditions the healthy (H), partially broken gear (PBG), and partially cracked teeth (PCT) a variety of signal processing techniques were applied to the vibration data to find any noticeable variations in the vibration patterns that would lead to gear failures. The signal processing techniques used in this research are time-domain analysis, frequency-domain analysis with the Fast Fourier Transform (FFT), and time-frequency analysis with the Short-Time Fourier Transform (STFT).

2.3.1. Time-Domain

Time-domain analysis was used to track variations in vibration amplitude over time and look for any obvious anomalies like spikes or irregularities that would point to gear failures. The vibration data time-domain signals were made for each of the three gear circumstances at each voltage levels as shown in Figure 4. At the time-domain approach there were no appreciable visual variations between the PCT, PBG, and Healthy sections and no anomaly in the data from healthy and defective gears was identified. This is due to the noise factor that the sensor encounters during vibration measurements.

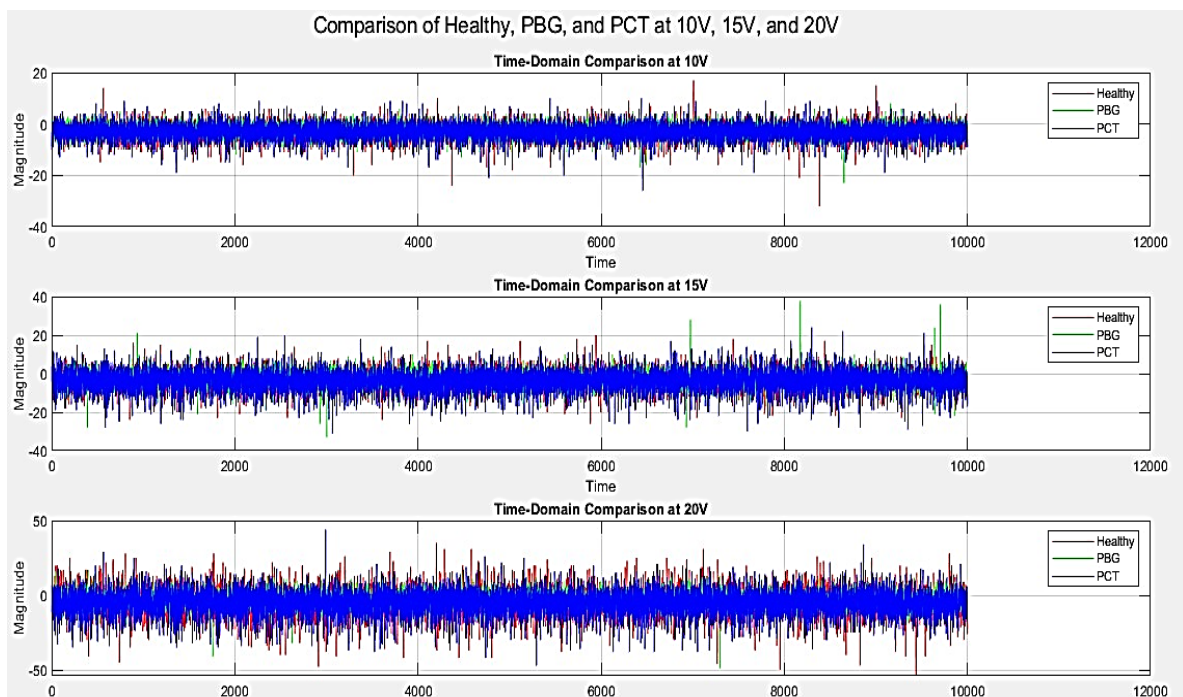


Fig. 4. Time-Domain Vibration Signals for Healthy, Partially Broken Gear, and Partially Cracked Teeth Conditions at 10V, 15V, and 20V

2.3.2. Frequency domain analysis

The Fast Fourier Transform (FFT) was performed on time-domain data to obtain additional understanding of the frequency properties of the vibration signals. By breaking down the time-domain signal into its frequency components, this transformation makes it possible to pinpoint prominent frequencies that may be connected to gear failures. For each of the three voltage levels the FFT was figured out and the frequency spectrum that was produced were shown for comparison as shown in Figure 5.

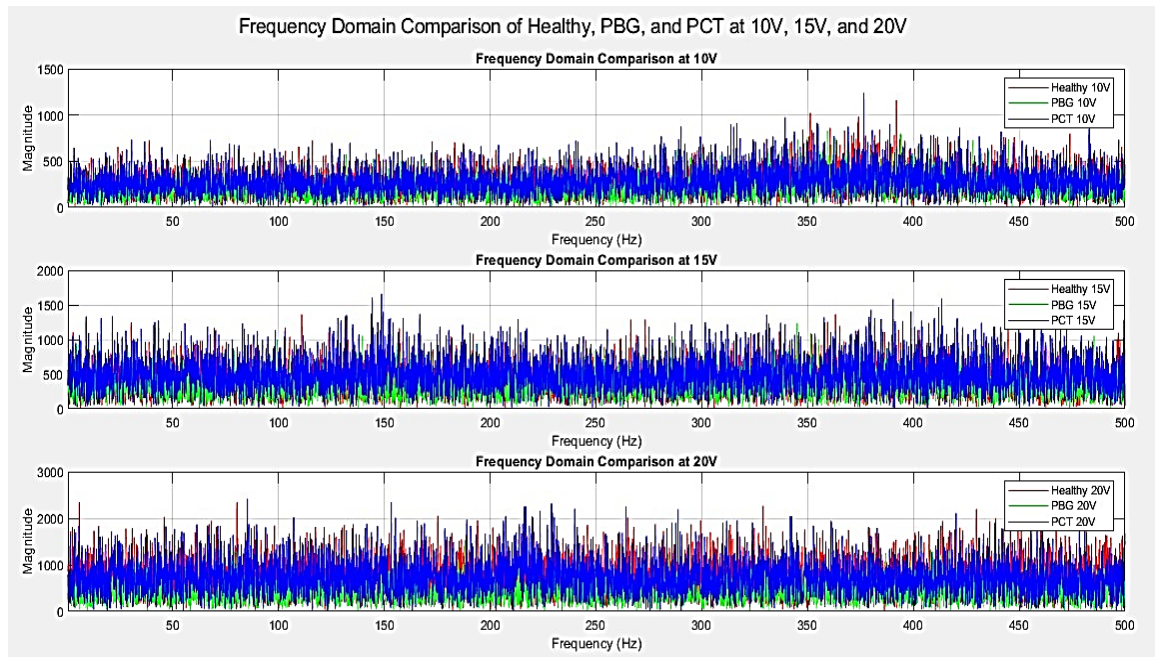


Fig. 5. Frequency-Domain Spectra of Vibration Signals for Healthy, Partially Broken Gear, and Partially Cracked Teeth Conditions at 10V, 15V, and 20V

As shown in Fig.6, same as for time-domain analysis, the frequency spectra did not show any significant differences between Healthy, PBG, and PCT conditions, which means a more advanced approach is required for accurate prediction of gears.

2.3.3. Fourier Transform in Short Time (STFT)

STFT was used to capture the time-varying nature of the frequency content in the vibration signals. It is possible to observe how the frequency components vary over time through STFT time frequency representation of the signal. Therefore, the behavior of signal in frequency and time domain can be checked. This is especially helpful in identifying intermittent defects or transient behavior that might not be seen in the whole frequency band. As shown in Figure 6 STFT results did not show any noticeable or consistent differences between the working and malfunctioning gears despite this thorough investigation and further confirming the subtlety of the faults.

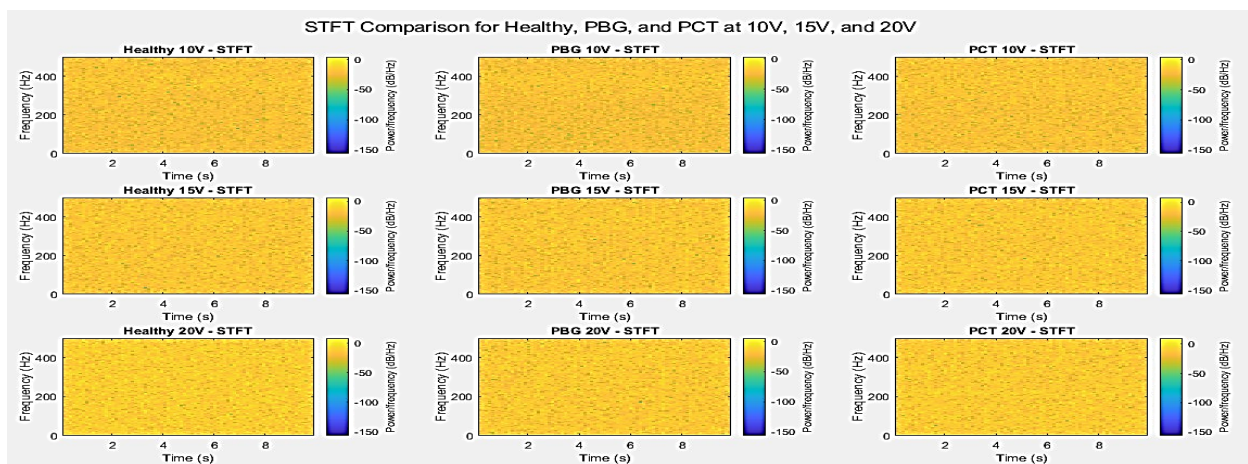


Fig. 6. STFT at each signal

2.3.4. Scalograms

The Scalograms, which are visual representations of the wavelet transform of a signal, are used to analyze time-frequency characteristics. They are generated when there is a need for time localization for short duration and high frequency events. The scalogram images are produced through applying wavelet transform to the signal. The scalograms of all the profiles are shown in Figure 7. In the current

case, it also does not show any prominent pattern which can be used as an identification to draw some useful conclusion regarding the profile of gear.

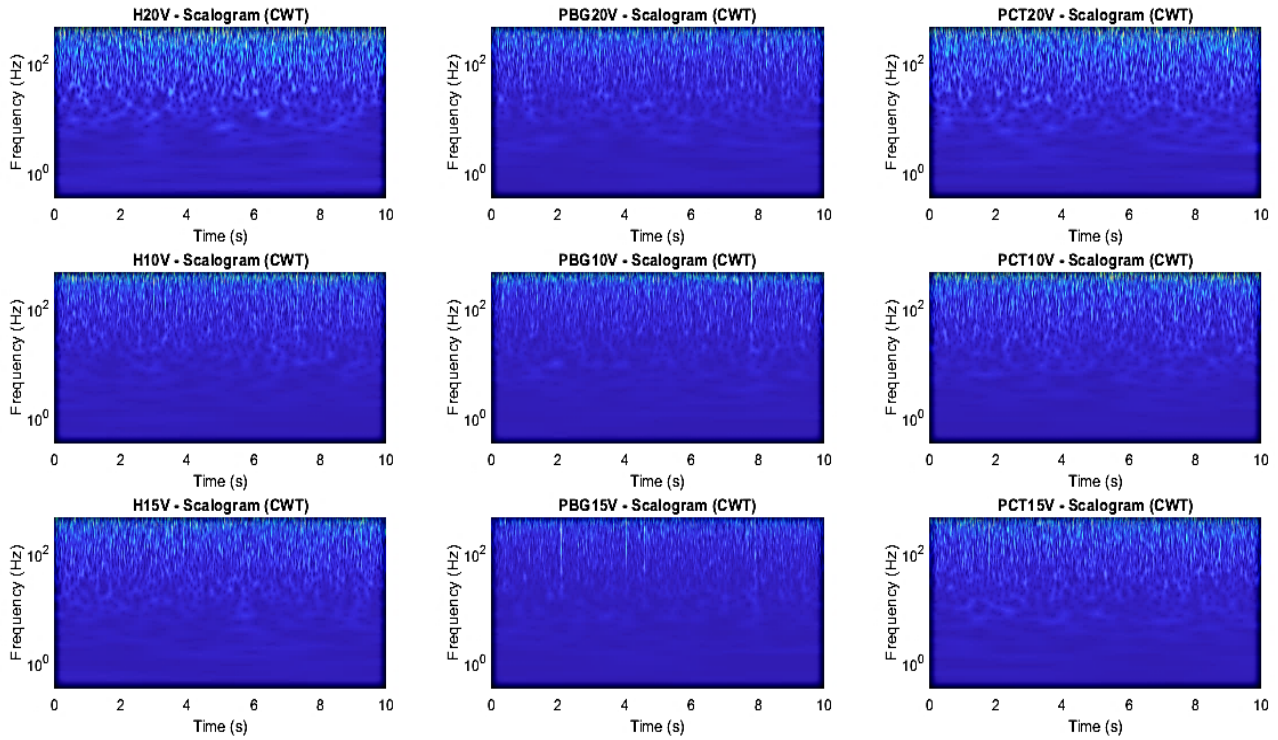


Fig. 7. Scalogram Images

The scalograms analytical methods showing the vibration signals from the healthy, partial broken gear, and partial crack teeth gears did not show any appreciable or distinct changes. The discovery implies that conventional signal processing techniques would not be enough in capturing the nuanced aspects of gear defects in this specific configuration. Consequently, machine learning techniques will be further investigated for improved fault identification.

3. Machine Learning Model Development

The model was trained through supervised learning. The steps involved are divided into the preparation of data and through supervised learning training of developed model through labeled data as shown in Figure 8. The cross validations of 10 are adopted in which the data is divided into 10 sections (folds) that means that on nine sections, the model is trained, and on the remaining part, it is tested [19]. Ten times through this process, one test set is used for each fold. By using this method, overfitting is prevented, a situation in which the model performs well with training data but badly with unknown data.

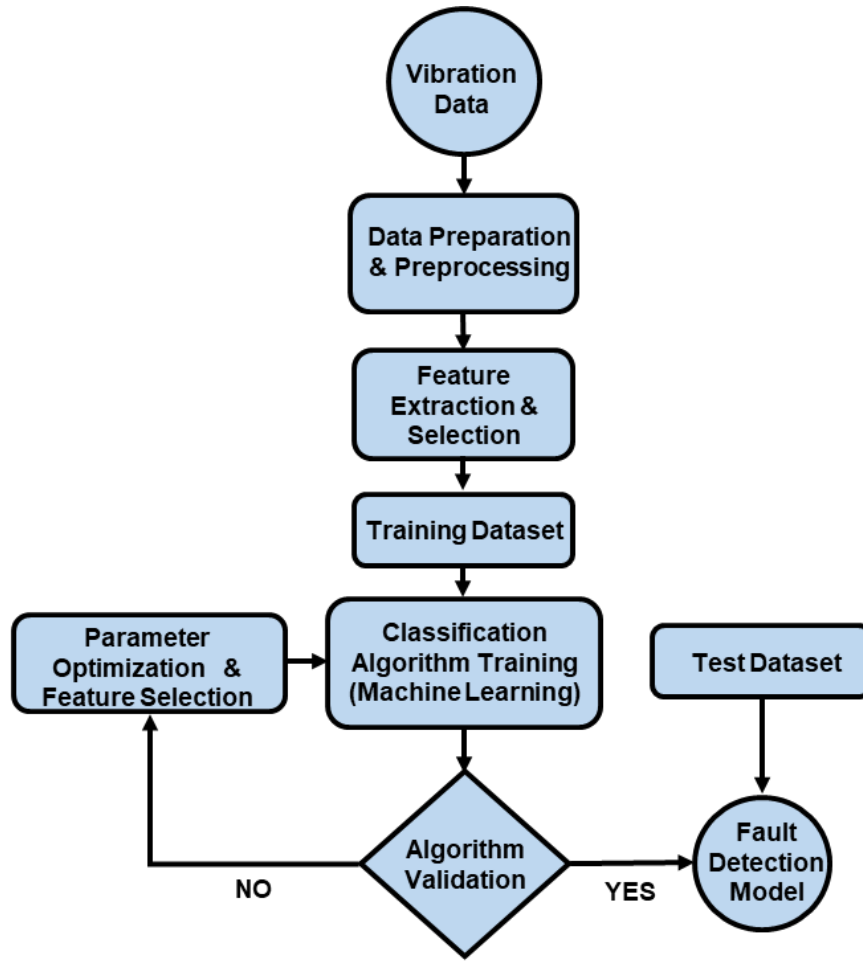


Fig. 8. ML Model workflow

3.1. Feature Extraction

A total of 15 features were extracted from the extracted signals that belongs to time-domain and frequency-domain. The time-domain features are calculated using statistical measures from the raw signal over time. It is used to capture useful characteristics of amplitude variations over time and are used in various domains, especially in fault diagnosis. There are a total of 15 features in time-domain and in spectral kurtosis shown in Table 1.

Table 1. Time Domain Features

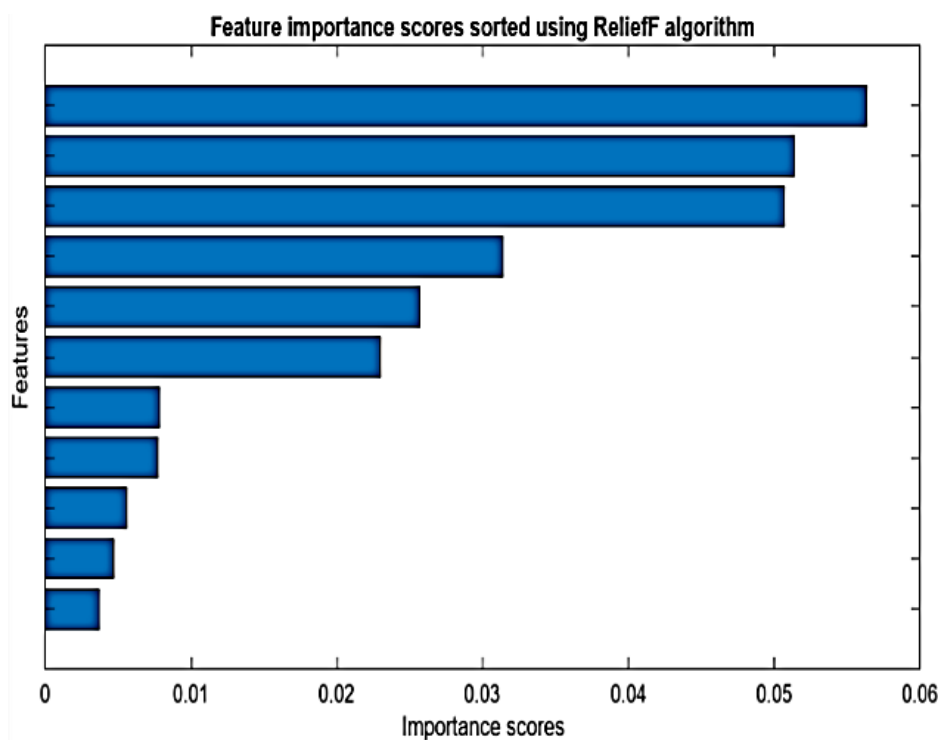
Time Domain Features	
Sr#	Features
1	Mean
2	Standard Deviation (Std)
3	Skewness
4	Kurtosis
5	Peak-to-Peak
6	Root Mean Square (RMS)
7	Crest Factor
8	Shape Factor
9	Impulse Factor
10	Margin Factor
11	Energy

Table 2. Spectral Kurtosis Features

Spectral Kurtosis	
Sr#	Features
1	Spectral Kurtosis Mean
2	Spectral Kurtosis Std
3	Spectral Kurtosis Skewness
4	Spectral Kurtosis Kurtosis

3.2. Feature Selection

There was a total of 15 features that were extracted shown in Table 1 so ReliefF algorithm used which is a feature selection method that ranks features based on their relevance to the target variable to identify and remove features with a negative impact that are identified and removed. So, after using ReliefF algorithm the results are shown in Figure 9.

**Fig. 9.** Selected Features Using ReliefF algorithm

The results illustrate the feature selection process using the ReliefF algorithm, which evaluates all 15 features shown in Table 1 and removes four with negative impact, resulting in an optimized set of 11 features for model training. Consequently, a total of 11 features were extracted and the model is trained based on these features to enhance the accuracy.

3.3. Results and Discussions

The model shows 72% accuracy through the Fine Tree Model. The accuracy of all the trained models is shown in Figure 10.

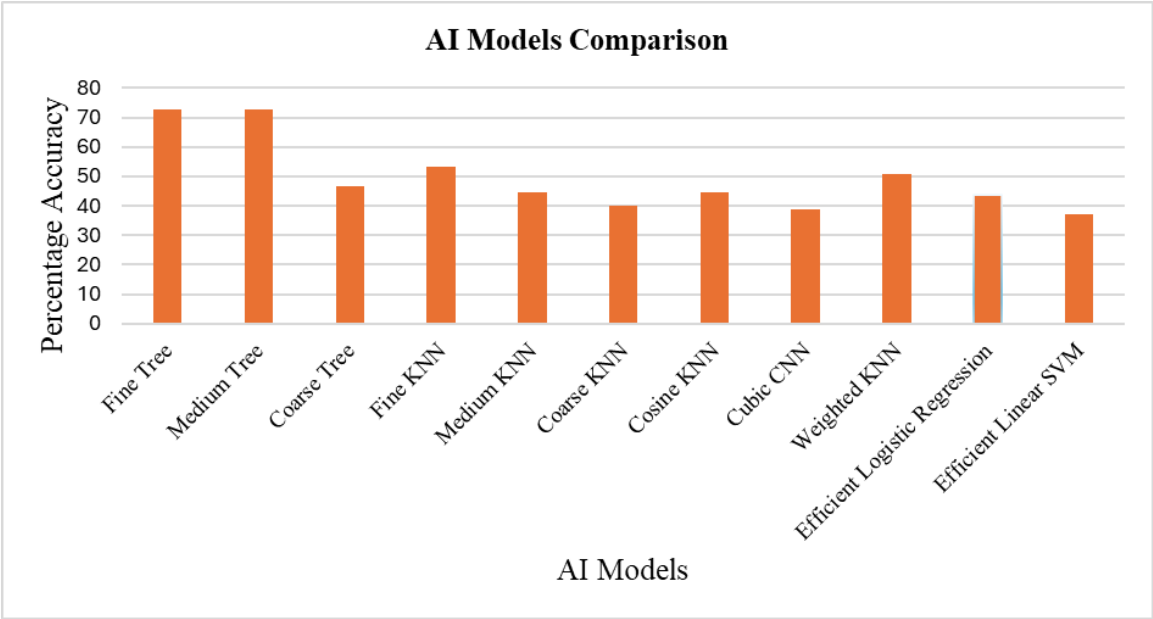


Fig. 10. ML Model Comparison

Through it can be identified, the accuracy of Fine Tree and Medium Tree AI models gives the highest accuracy, generating overall accuracy of 72%. The confusion matrix is shown in Figure 11.

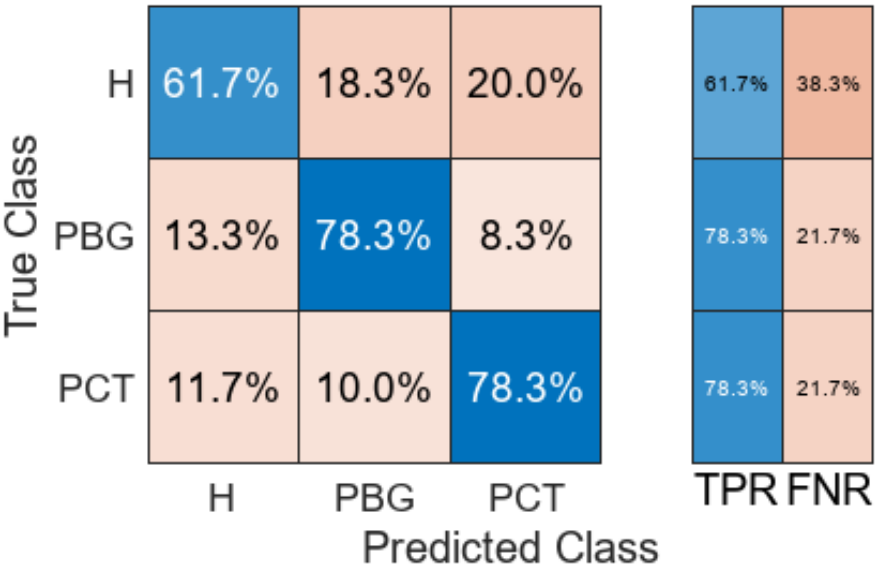


Fig. 11. Confusion matrix

The trained model can predict the Healthy state with an accuracy of 61% and have a high ability to predict the defective two cases to be 78.3%. The Receiver Operating Characteristic (ROC) curve is plotted in which the relation is between the true positive rate (sensitivity) versus the false positive rate for various categorization thresholds. It gives an indication of how well the model can distinguish between the various classes. Area Under the ROC Curve (AUC) summarizes the model ability to distinguish between positive and negative classes in a single numerical value. Excellent performance is indicated by a model with an AUC near 1, whereas performance that is no better than random guessing is represented by an AUC of 0.5. The ROC curve of trained model is shown in Figure 12.

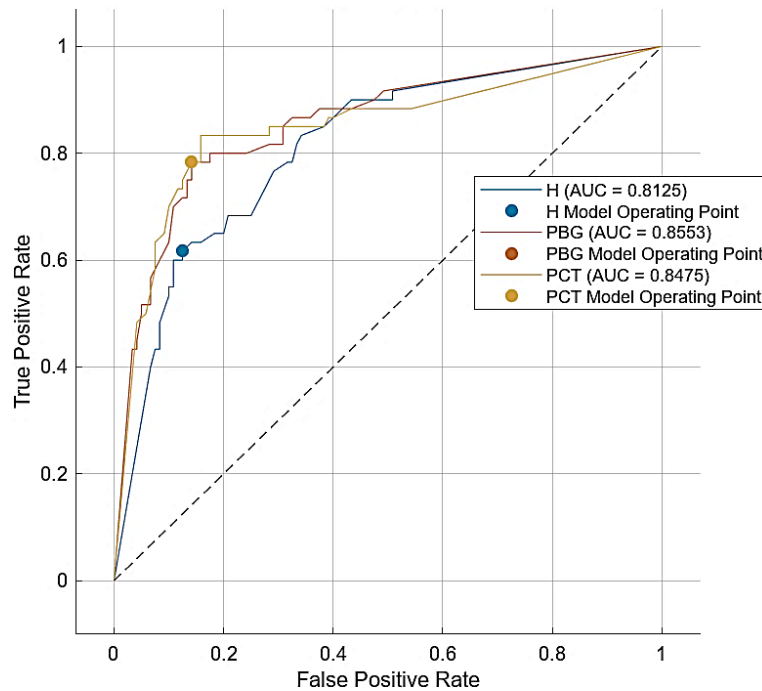


Fig. 12. ROC Curve

It shows that the AUC to be closed to one for the defective two cases and around 0.81 for the Healthy state. The above results show the accuracy of the model which can be implemented to any system to figure out the health assessment of spur gears and in the later stage able to predict the maintenance of spur gears through statistical analysis of the required-results. The result produced showed the Tree model to have most accuracy compared to all the other models.

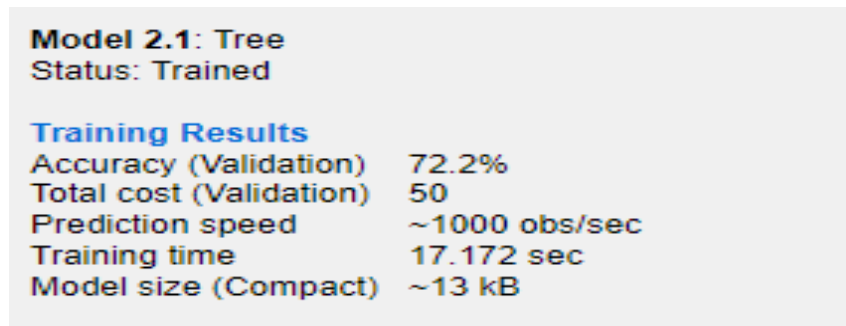


Fig. 13. Trained Model Result.

The Tree model was successfully trained and validated giving validation accuracy of 72.2% showing that nearly this percent of prediction made by model validation were correct. The total validation cost is shown in Figure 13 in which the total cost value of 50 refers to the misclassification cost or loss function value during validation. It basically indicates a predefined penalty on the model performance during validation. Also, the model has the capability of predicting 1000 observations per second. The approximately time of training is 17.172 seconds showing that it is trained relatively faster. The model's weight is approximately 13kB, showing a lightweight mode.

The validation accuracy of 72% achieved by the Fine Tree Model exceeds the target of 70%, which was selected as a practical benchmark based on the complexity of the vibration signals and the computational limitations of real-time monitoring systems. Achieving higher accuracy, such as 80%, would require more advanced feature extraction techniques or deeper models, which could increase computational demands and make the system less feasible for industrial deployment. Conversely, an accuracy below 70% would likely result in unreliable fault detection, as observed in prior studies [20], leading to increased false positives or negatives. So, based on this, the overall performance is good, suitable for real-time monitoring and useful where computational efficiency is of priority.

4. Conclusion

The main objective of this study was to use an accelerometer to collect data from the gear meshing profile, which was well accomplished. Many signals processing methods, including time-domain analysis, frequency-domain analysis, Short-Time Fourier Transform (STFT), and scalograms, were applied to the raw vibration data. Even though, the signal processing techniques were unable to produce patterns that might be used to discern between the various states of the equipment. This made a more sophisticated method necessary. These limits were addressed using machine learning, which developed a supervised learning model with an initial accuracy of 68% by extracting 15 important features from the raw signal. Eleven optimized features remained after more refining using the Relief algorithm eliminated any characteristics that had a detrimental effect or were superfluous. As a result, the accuracy of the model increased to 72%. The Tree model was proposed giving the highest accuracy compared to others adopting the current scheme of work. The model presented in this work is a potential real-time gear condition monitoring solution, and its application to Industry 4.0 contexts for predictive maintenance could be expanded. Additionally, this work shows that feature refinement can improve the accuracy of machine learning-based models. Subsequent investigations may investigate the utilization of deep learning methodologies, such as those proposed by[21], to enhance the precision and versatility of the model on increasingly intricate datasets. While the model achieved a validation accuracy of 72%, surpassing the 70% target, future recommendations could explore advanced deep learning techniques to push the accuracy closer to 80%, potentially at the cost of increased computational resources.

References

- [1] Nazarova Nafisa Ilhomovna., "The Importance of role of Gears in Mechanical Engineering," in Proceedings of the Thirteenth EuroSys Conference, Porto Portugal: ACM, Apr. 2018, pp. 1–15. doi: 10.1145/3190508.3190551
- [2] J. A. Faria, E. Nunes, and M. A. Matos, "Cost and quality of service analysis of production systems based on the cumulative downtime," *Int. J. Prod. Res.*, vol. 48, no. 6, pp. 1653–1684, Mar. 2010, doi: 10.1080/00207540802068631.
- [3] E. P. Carden and P. Fanning, "Vibration based condition monitoring: A review," 2004. doi: 10.1177/1475921704047500.
- [4] N. H. Chandra and A. S. Sekhar, "Fault detection in rotor bearing systems using time frequency techniques," *Mech Syst Signal Process*, vol. 72–73, pp. 105–133, May 2016, doi: 10.1016/j.ymssp.2015.11.013.
- [5] T. Praveenkumar, M. Saimurugan, P. Krishnakumar, and K. I. Ramachandran, "Fault diagnosis of automobile gearbox based on machine learning techniques," in *Procedia Engineering*, Elsevier Ltd, 2014, pp. 2092–2098. doi: 10.1016/j.proeng.2014.12.452.
- [6] Y. Bella, A. Oulmane, and M. Mostefai, "Industrial Bearing Fault Detection Using Time-Frequency Analysis," 2018. [Online]. Available: www.etasr.com
- [7] P. Ong, A. John Koshy, K. H. Lai, C. K. Sia, and M. Ismon, "A deep learning approach for health monitoring in rotating machineries using vibrations and thermal features," *Decision Analytics Journal*, vol. 10, Mar. 2024, doi: 10.1016/j.dajour.2024.100399.
- [8] M. Sohaib et al., "Gearbox fault diagnosis using improved feature representation and multitask learning," *Front Energy Res*, vol. 10, Sep. 2022, doi: 10.3389/fenrg.2022.998760.
- [9] H. Sindi, M. Nour, M. Rawa, Ş. Öztürk, and K. Polat, "A novel hybrid deep learning approach including combination of 1D power signals and 2D signal images for power quality disturbance classification," *Expert Syst Appl*, vol. 174, Jul. 2021, doi: 10.1016/j.eswa.2021.114785.

-
- [10] F. Tan, “Improving Feature Selection Techniques for Machine Learning Improving Feature Selection Techniques for Machine Learning Feng Tan”, doi: 10.57709/1059437.
 - [11] F. Tan, “Improving Feature Selection Techniques for Machine Learning Improving Feature Selection Techniques for Machine Learning Feng Tan”, doi: 10.57709/1059437.
 - [12] M. A. Khamis, W. Gomaa, and W. F. Ahmed, “Machine learning in computational docking,” Mar. 01, 2015, Elsevier. doi: 10.1016/j.artmed.2015.02.002.
 - [13] F. P. García Márquez, A. M. Tobias, J. M. Pinar Pérez, and M. Papaelias, “Condition monitoring of wind turbines: Techniques and methods,” *Renew Energy*, vol. 46, pp. 169–178, Oct. 2012, doi: 10.1016/j.renene.2012.03.003.
 - [14] P. Nunes, J. Santos, and E. Rocha, “Challenges in predictive maintenance – A review,” Feb. 01, 2023, Elsevier Ltd. doi: 10.1016/j.cirpj.2022.11.004.
 - [15] H. Bhavsar and A. Ganatra, “A Comparative Study of Training Algorithms for Supervised Machine Learning,” vol. 2, no. 4, 2012.
 - [16] L. Lévesque, “Revisiting the Nyquist criterion and aliasing in data analysis,” *Eur. J. Phys.*, vol. 22, no. 2, pp. 127–132, Mar. 2001, doi: 10.1088/0143-0807/22/2/304.
 - [17] Y. Yu et al., “Dynamic control flow in large-scale machine learning,” in *Proceedings of the Thirteenth EuroSys Conference*, Porto Portugal: ACM, Apr. 2018, pp. 1–15. doi: 10.1145/3190508.3190551.
 - [18] I.O. Muraina, “Ideal Dataset Splitting Ratios in Machine Learning Algorithms: General Concerns for Data Scientists and Data Analysts”.
 - [19] C. Schaffer and S. A. H. C. Edu, “Selecting a Classification Method by Cross-Validation,” 1993.
 - [20] B. Al-Najjar, “Impact of real-time measurements of operating conditions on effectiveness and accuracy of vibration-based maintenance policy - A case study in paper mill,” *J. Qual. Maint. Eng.*, vol. 6, no. 4, pp. 275–287, Dec. 2000, doi: 10.1108/13552510010346815.
 - [21] A. A. Rasheed, “Improving prediction efficiency by revolutionary machine learning models,” *Mater. Today Proc.*, vol. 81, pp. 577–583, 2023, doi: 10.1016/j.matpr.2021.04.014.