SLJoAS Sri Lankan Journal of Applied Sciences

Research Article

www.sljoas.uwu.ac.lk/index.php/sljoas

Development of Fault Detection and Diagnosis Model for Drilling Machines

A.T.C. Dhivya *, R.M.T.C.B. Ekanayake, D.D.B. Senanayake
Department of Science and Technology, Faculty of Applied Sciences, Uva Wellassa University of Sri Lanka, Passara Road, Badulla, Sri Lanka, 90000

*** Corresponding author email address**: sct19050@std.uwu.ac.lk
(Received 22nd November 2024; Accepted 24th April 2025)

Abstract

Drilling machines are essential in industrial applications as they are used to drill materials such as metal, plastic, and concrete and are now being incorporated into smart industries. Such machinery needs to be maintained properly given that they are known to wear out very quickly. In its previous form, as a manual process, monitoring has served its purpose significantly. Nowadays, it is replaced by automated systems that utilize achievements in signal processing and machine learning. This work proposes fault detection for drilling machines through sound signals and Fine K Nearest Neighbour (Fine KNN). Fine KNN was selected due to its moderate accuracy and computational efficiency compared to other classifiers in real-time despite a slightly lower accuracy than Quadratic SVM or bagged trees. The dataset employed is obtained from a GitHub repository and contains sound signals under various fault scenarios: healthy, bearing, gear, and fan. In feature extraction, 16 time-domain and frequency-domain features are extracted from the chosen signal and then narrowed down to 12 by using the RelieF algorithm to improve the model. The Fine KNN model maintains an efficiency of operation while detecting faults at a rate of 95.6%, which indicates the model's accuracy. For this reason, feature selection and preprocessing serve a critical role in enhancing the model performance and suitability for real-time applications as affirmed by this research. Thus, this research opens up the possibility of integrating more complex models for machine condition monitoring at the edge devices. Future work will focus on obtaining more sophisticated classifiers and better preprocessing for improving fault detection performances in compact and power-efficient platforms suitable for the industrial IoT environment.

Keywords: Fault detection, Fine KNN, Sound Signals, Monitoring, Accuracy

1. Introduction

Drilling machines are fundamental in the industry and play a pivotal role in various applications. The main functionality of the drilling machines is simple but it is mandatory for a vast array of operations that are essential to the production of machinery, infrastructure, construction, and manufacturing. Drilling machines, while being vital, are also subject to mechanical wear and tear, which can result in downtime or operational failures if not adequately maintained. Since most industries depend heavily on these machines for several basic operations and also for efficient production, ensuring their reliability and longevity is of utmost importance. This has led to the increasing importance of machine condition monitoring and fault detection systems.

In the past, methods such as manual sound monitoring by experienced operators were used to detect faults in machines through sound, vibration, or visual inspection. While this technique was useful, it heavily relied on human expertise, including the operator's experience, variability in human perception, hearing ability, and concentration, which lacked consistency. According to Brito et al. [1]An unsupervised approach for fault detection is highly crucial in industrial machinery.

This research aims to find a solution to find anomalies in the drilling machines while addressing the following objectives, to investigate the feasibility of utilizing sound signals for fault detection in drilling machines, to develop a robust machine learning model for fault detection and diagnosis in drilling machines using sound signal analysis, to evaluate the performance of different edge machine learning algorithms for fault detection in drilling machines, to compare the results with existing fault detection methods to assess the effectiveness of the proposed approach. Edge machine learning refers to implementing and test the machine learning algorithms directly on the edge devices like sensors or IoT devices, which are directly in contact with the system instead of sending it to cloud computing. This method enables quick response and is beneficial for real-time monitoring. Achieving these objectives will develop a model to deploy on an edge device for real-time fault detection.

2. Literature Review

In the past few years, there has been great development in the ways of carrying out the automation of mechanical fault diagnosis and inspection. The origin of the problem of fault detection and diagnosis goes back to early industrial systems in which machines were primarily inspected through human judgment and operator experience. This method improved over time along with the emergence of better systems. The first steps of the development of such automatic fault detection methods can be traced back to the nineties of the last century along with the inventions of vibration analysis and signal-related provision. Development of this direction paved the way for combinations of sensors and data acquisition systems which helped enhance remote monitoring and more precise detection of faults.

The examination of acoustic signals as a diagnostic tool has become popular through its ability for remote monitoring. Altaf et al. [2] Pointed out the limitation of using vibration sensors on unreachable machines and suggested the classification of faults through audible signals, working with such methods as Kernel Linear Discriminant Analysis (KLDA), Support Vector Machines (SVM), and k-Nearest Neighbours (KNN). Same thing, Hongmei Liu, Lianfeng Li and Jian Ma[3] Performed sound fault diagnosis without time-consuming procedures concerning feature selection utilizing deep learning. In addition, the researchers noted the limitations of such approaches in terms of timely engagement due to the heavy requirements of spectrogram construction. However, in both research papers, the authors focused solely on the bearing fault and failed to broaden their focus area.

To the above Shubita, Alsadeh, and Khater [4] Successfully forecasted faults at an early stage employing acoustic emission (AE) and reported 96.1% accuracy using a fine decision tree, highlighting the need for real-time monitoring via IoT devices. Their methodology closely aligns with the methodology used in the present study. Based on their comparative analysis, they opted for the Fine Decision Tree model for deployment because of its low computational overhead. However, this type of model is highly prone to overfitting, especially in the presence of noisy real-world data. In contrast, Fine KNN offers better generalization by considering neighborhood-based class distributions. Also, Senanayaka et al. [5] Improved acoustic signal processing with DEMUCS and 1D-CNN allowing effective signal faulting isolation and thus efficacy beyond blind source separation techniques. But this method requires high computational efficiency which is not suitable for edge devices.

Many researchers have worked with vibration analysis techniques for fault evaluation. Swapnil K. Gundewar and Prasad V. Kane [6] Examined bearing in terms of breakdown by applying denoised vibration signals and neural network with an accuracy of 99.58%, but have not fully implemented on edge devices. The use of high-end hardware in Gundewar and Kane's study renders it unsuitable for real-time fault diagnosis on edge devices. Khalil and Rostam [7] Worked on a semi-automated vibration diagnostic method by applying Fast Fourier Transform (FFT) and ensemble ML models enabling them to provide enhanced early fault detection. Brandao and Costa [8] Also worked on feature-based misalignment faults, successfully extracted using Fast Fourier Transform's technique with Support Vector Machine being the best of the other classifiers. These works center around the vibration signals whereas the present study is based on acoustic signal analysis as a cost-effective, flexible, and non-invasive alternative, making it more suitable for environments where sensor placement for vibration monitoring is impractical.

Tran, Pham and Lundgren [9], on the other hand, focused on one of the most prevalent issues which is the lack of balance in the datasets in drill failure detection, and used CNN with Long Short-Term Memory LSTM and attention mechanism to attain an overall accuracy of 92.35%. Vununu et al. [10] Proposed a novel deep convolutional autoencoder (DCAE) architecture for power spectrum density image feature extraction that was effective in highly noisy environments. These approaches demonstrate the potential of deep learning in fault detection, yet there is a gap in developing efficient, interpretable, and real-time solutions that are capable of running under resource constraints.

These were however not enough to make the last models stand-alone without limitations Hybrid methods have come up to remedy these traditional limitations. Kiran Vernekar and KV Gangadharan [11] Utilized neural networks and decision trees simultaneously for gearbox fault diagnosis and achieved an overall accuracy of 85.5% yet still managing the challenging dynamics of such machines. Liu et al. [12] Used a dynamic unscented Kalman filter to effectively utilize computational capacity on a rotary steerable drilling tool, lowering both the number of missed alarms and the amount of computation time.

Jonguk Lee et al. [13] Extracted Mel-frequency cepstral coefficients (MFCCs) from the audio signals and also employed SVM for classification. Their dataset, which was gathered from an NS-AM-style railway point machine at Sehwa Company in Daejeon, South Korea, had an accuracy of 94.1 percent. Every sound in their dataset had a duration of roughly 5000 milliseconds. However, because each sound recording is so brief, their method did not yield a promising outcome when applied to our drill sound collection.

3. Methodology

In Figure 1, a schematic illustration of the research procedures is outlined. It starts with the audio recording made by a mobile phone placed 10 cm away from the testing instrument, a drilling machine. After this, the recorded audio

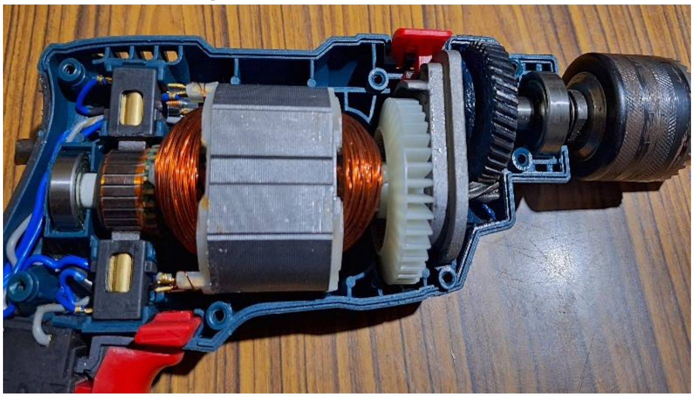


Fig.1. Internal components of a drilling machine

signals are sent to a host PC with MATLAB for further processing. Further processing steps performed within the MATLAB software are depicted in Figure 2, highlighting the key stages and their interconnections.

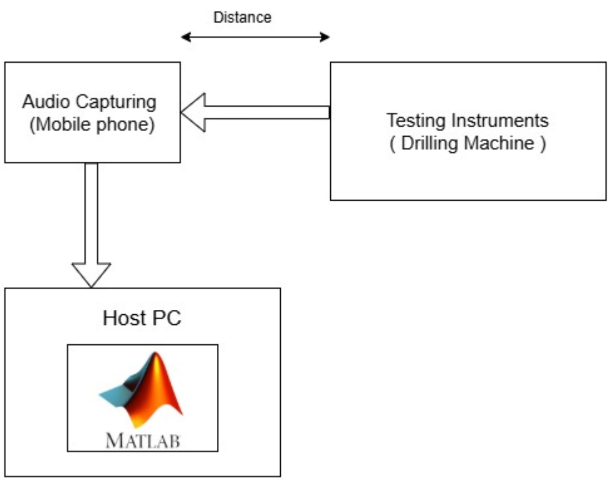


Fig.2. Schematic overview

Fig.3. Workflow

3.1. Data acquisition

In this research, machine sound signals were classified into four categories: normal state, bearing fault, gears fault, and fan fault. These sound signals were obtained from varying states of a drill machine, employing a smartphone as a data acquisition tool. The data used in this study was obtained from an online repository made available on GitHub [14] containing sound signals under different fault conditions and was used as such without further data collection and pre-processing since collecting accurate data was difficult. According to the information provided on GitHub, the smartphone was placed approximately 10 cm away for recording, and the CROWN power tool (CT10128 drill) was used for data collection. Each of the extracted features has 7115 values, and there is the same number of samples for each of the four fault conditions—normal, bearing gear, and fan. Balanced samples like this ensured impartial training and testing of machine learning models. This secondary data was feature extracted, and used in training the machine learning model for diagnosing faults in MATLAB R2023a. [15]

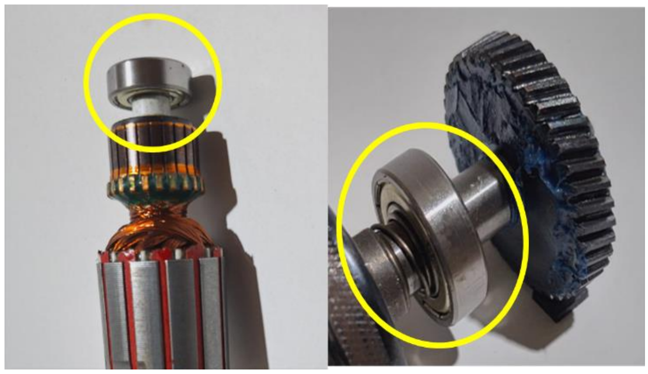


Fig.4. Bearing

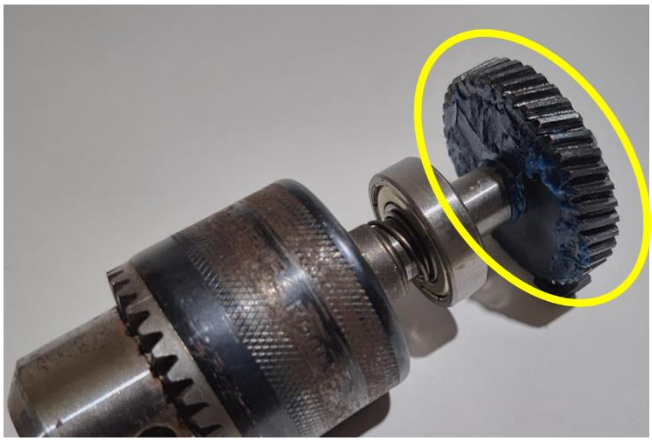


Fig.5. Gear

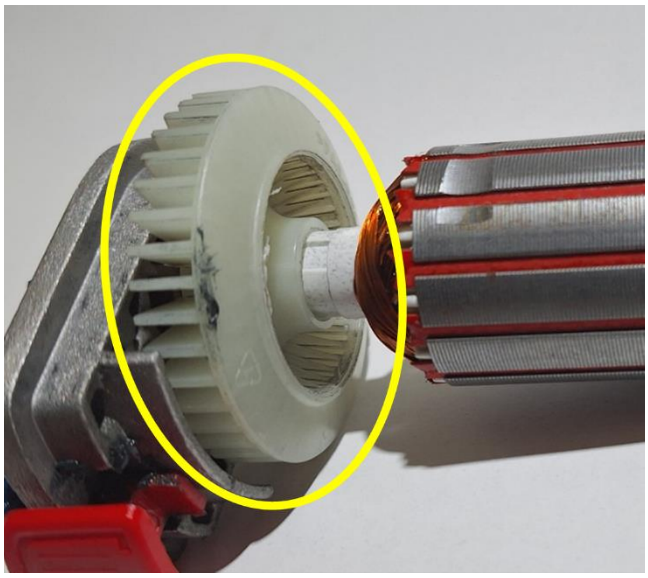


Fig.6. Fan

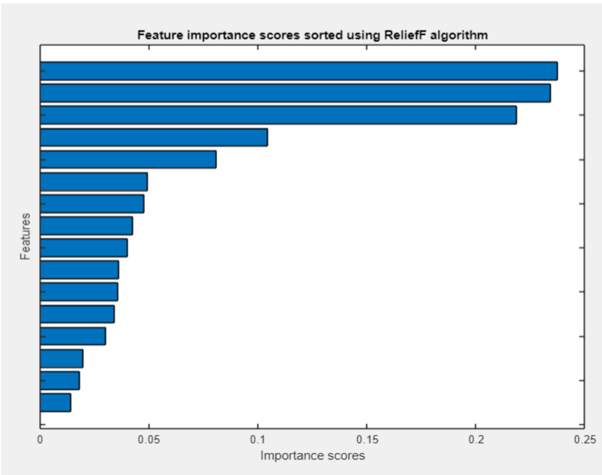
3.2. Data preprocessing

Preprocessing is one of the most important steps in the detection of faults using sound signals as a method since this is where raw data is fine-tuned in a way that eliminates as much noise as possible for the next analysis step. In this research, the pre-processing began with the Hanning window, especially for unknown signals, the results of the fast Fourier Transform (FFT) are more realistic. The choice of the right window size is important because it determines the accuracy of the classification. The trade-off window size of 2048 points was selected to have an optimal trade between the analysis accuracy and memory limitation, and the 75 % overlap was used to improve consideration evaluation. Furthermore, a digital bandpass filter was used to eliminate the DC component from the sound signals using the cutoff frequencies of 20Hz and 20kHz as the anti-aliasing filter[16]. These steps aided in improving the quality of the signal to enhance feature extraction and to diagnose the fault while training the machine learning model in MATLAB.

3.3. Feature evaluation

A total of 16 features were extracted for each data window, with 10 features coming from the time domain and 6 features from the frequency domain. Features extracted from the time domain are, RMS, Mean, Median, Variance, Skewness, Kurtosis, Shape factor, Crest factor, Impulse factor, and Margin factor, and from the frequency domain are Peak 1, Peak 2, Peak 3, PeakLocs1, PeakLocs2, and PeakLocs3. As a next step, extracted features were evaluated to identify the impact value of each feature on the model accuracy. For that, features were ranked using the RelieF algorithm.

According to Figure 7 below, Peak 1 is the most important feature, followed by RMS, Variance, and Peak 3



	Select	Features	ReliefF
1	✓	Peak1	0.2374
2	✓	RMS	0.2342
3	✓	Variance	0.2188
4	✓	Peak3	0.1045
5	✓	Peak2	0.0808
6	✓	PeakLocs1	0.0494
7	✓	ImpulseFactor	0.0478
8	✓	CrestFactor	0.0425
9	✓	Kurtosis	0.0402
10	✓	ShapeFactor	0.0362
11	✓	PeakLocs3	0.0359
12	✓	PeakLocs2	0.0340
13	✓	Median	0.0302
14	✓	Mean	0.0196
15	✓	Skewness	0.0182
16	✓	MarginFactor	0.0140

Fig.7. Features ranking using RelieF Algorithm

Figures 8 and 9 depict the difference in the amplitude between sound signals of the drilling machines with no defect and with a bearing defect.

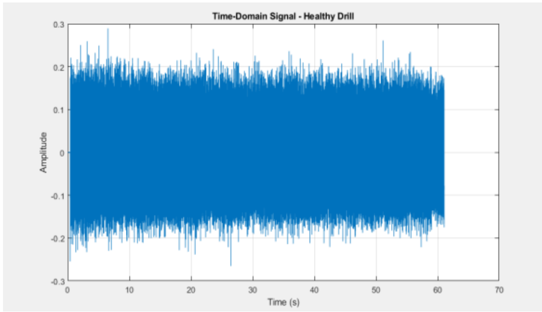


Fig.8. Time Domain signal for drill with no defects

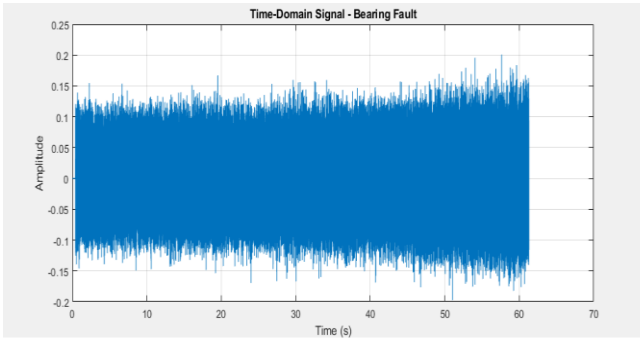


Fig.9. Time Domain signal for drill with defect in bearing

The figures below show histograms for Peak 1, Variance, Median and skewness for all the fault classes.

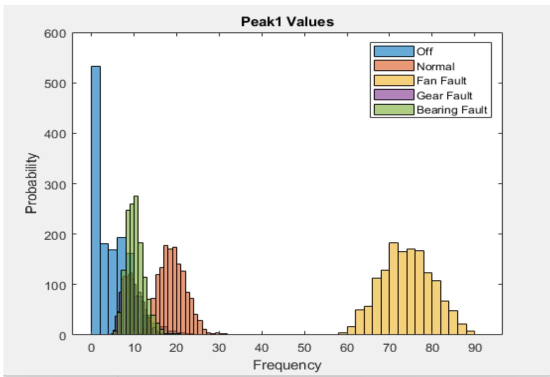


Fig.10. Histogram for Peak 1

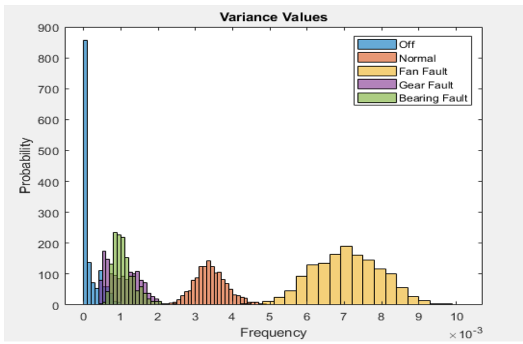


Fig.11. Histogram for Variance

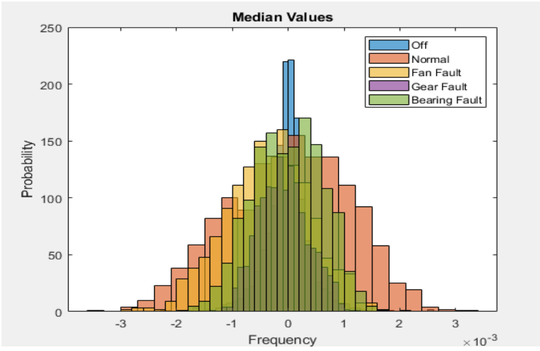


Fig.12. Histogram for median

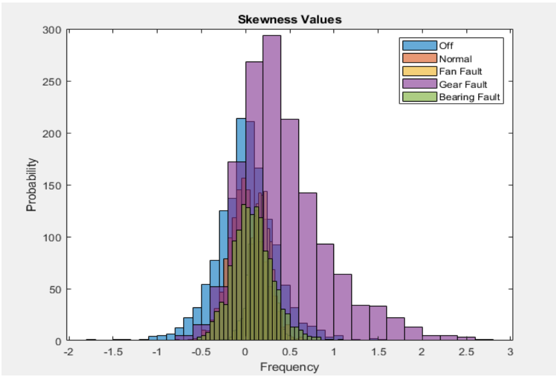


Fig.13. Histogram for Skewness

As shown in the Figures above, each fault is represented by a different color. A histogram plot is used to verify that the faults are ranked correctly. If the features are not overlapped, as in variance and Peak 1, then they represent favorable features that will help the model in classifying the faults. For insignificant features, like the median and skewness, the faults are overlapped in the histogram.

3.4. Model training

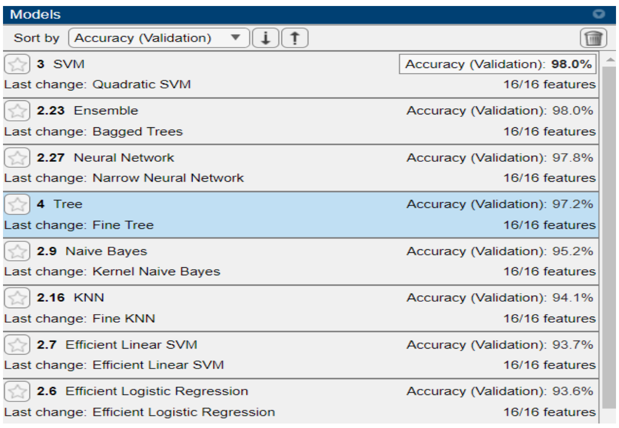


Fig.14. Highest model accuracies obtained

Following the ranking stage, the selected features are prepared for use with machine learning models. Various configurations and feature sets are tested with different ML techniques. Due to the relatively small size of the dataset, a cross-validation trade-off with k=5 is used. The highest precision is achieved by the quadratic SVM, which attained an accuracy of 98.0%, followed by the bagged trees ensemble classifier at 98%, the Narrow Neural Network at 97.8%, the fine decision tree at 97.2%, the Naïve Bayes classifier at 95.2%, and the Fine KNN algorithm at 94.1%.

The efficiency of the classification models is assessed using a confusion matrix, as illustrated in below Figure 15 where the column denotes the predicted class and the raw represents the true class. As a result, samples that are off-diagonal are misclassified, and samples that are diagonal are correctly classified.

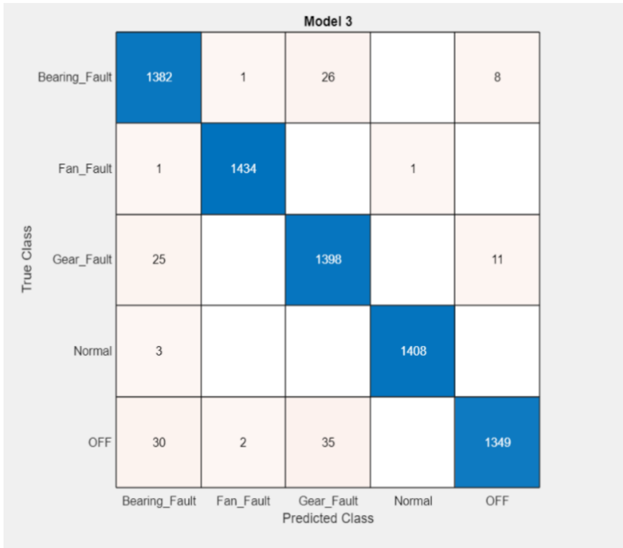


Fig.15. Confusion matrix for Quadratic SVM model

3.5. Model testing and validation

Model testing and validation is an essential process to show that produced FDD systems are reliable and accurate. From extracted feature values, 80% of the data was used for training, and the remaining 20% was used for testing. The model demonstrated accurate results while testing. However, as the CROWN power drill for which data was gathered was unavailable, real-time experimental validation in the laboratory was not possible. This is identified as a future scope of the work. In the training phase, Quadratic SVM and bagged ensemble classifiers showed good classification accuracy. On the other hand, for testing and prediction Fine KNN model was selected as it provides a better trade-off between simplicity and performance. A Fine KNN was chosen over Quadratic SVM or ensemble bagged classifiers because of its lower computational complexity and quicker inference time which are vital for real-time performance on embedded hardware. Although Quadratic SVM and ensemble bagged classifiers showed higher accuracy, they are computationally expensive so not suitable for real-time applications, especially on resource-limited platforms.

Specifically, Fine KNN provided a solution sufficiently accurate and effective enough to allow the developed fault diagnosis system to run on an embedded device. The proposed model was fine-tuned and validated on independently sampled data and model validation justified its stability and accuracy. The fine KNN model ensured high and stable accuracy overall while focusing on different faults such as bearing, fan, and gear, which makes it fit for real-time monitoring of machines.

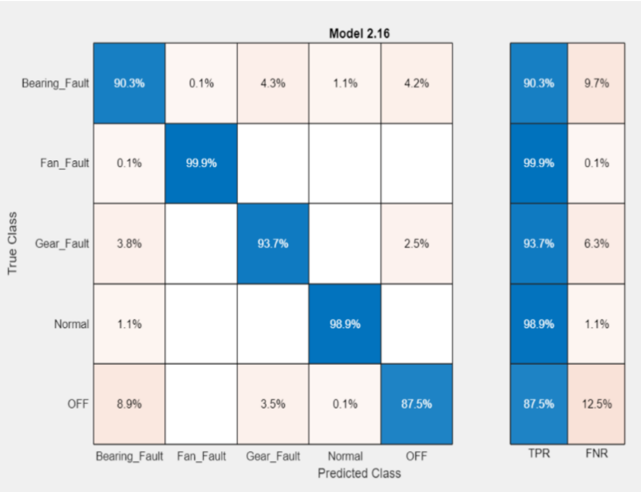


Fig.16. Confusion matrix for Fine KNN model

In the below ROC Curve, the True Positive Rate (sensitivity) is plotted against the False Positive Rate, and each line represents a different fault or state. Each fault type has a corresponding Area Under the Curve (AUC) value, which measures the model's ability to distinguish between classes. Higher AUC values, close to 1, indicate better performance. According to this principle, the Fan fault has the highest AUC while the bearing fault has a slightly lower AUC value. Normal and gear fault also have high AUC values depicting a good distinguishability.

The MATLAB codes used for all functions are available from the authors upon reasonable request.

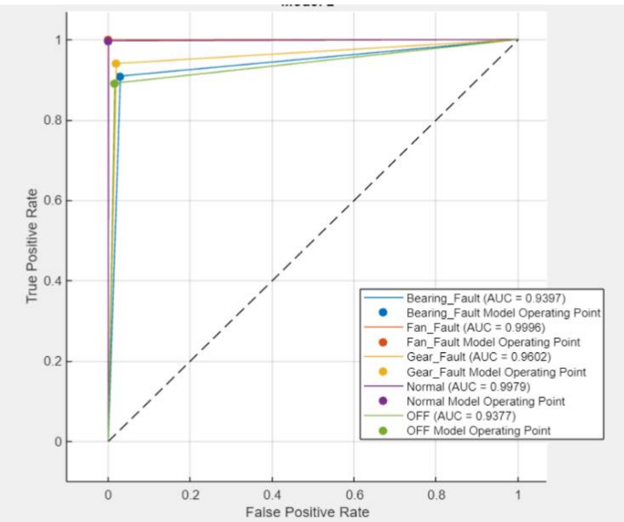


Fig.17. ROC curve for Fine KNN model

4. Results

The model testing and validation provided the reliability of the developed models and depicted each of its ability to be deployed on edge devices.

In the context of developing a fault detection model for drilling machines, the fine KNN model was selected based on a set of criteria that emphasized its effectiveness and suitability for deployment. This selection process was guided by the following key considerations:

- Accuracy: The fine KNN model was characterized by high classification accuracy in different types of faults, which is important for reliable fault detection.
- Computational Efficiency: Due to its low computational complexity, it showed shorter inference times which can be crucial for real-world applications on IoT devices.
- Robustness: The model showed comparatively steady outputs when the parameters were given mixed data sets, which makes it viable for real-life use.
- Scalability: This fine KNN model has exhibited the ability to scale up with increased data size which would make it relevant for complex fault detection tasks.
- Ease of Implementation: Because of its simplicity of concept, it can be easily integrated with other systems, which is useful when it comes to deploying it on edge devices.
- Feature Sensitivity: The model proved a way of using features that were extracted from audio data which was crucial in diagnosing other fault conditions such as bearings fans and gears.
- Validation Performance: Cross-validation outcomes provided evidence of the model's reliability and its ability to be generalized towards 'real-world' practice settings, increasing the safety of the model.
- Real-time Capability: Since the fine KNN model works perfectly in a real-time environment, the real-time detection and diagnosis of faults is feasible for monitoring the operation.

This model is mostly suitable for well-labeled data and is commonly used for classification and regression tasks. For the prediction of the value of new data points, KNN looks at the k closest (most similar) data points in the training data.

Although Quadratic SVM and ensemble bagged classifiers were marginally more accurate, Fine KNN was selected on theoretical grounds of computational efficiency. Fine KNN has lower training complexity, as it is a lazy learner and does not create a model during training. This reduces the front-end computational overhead, and it is thus preferred when speed of deployment or retraining must be prioritized.[17] In real-time fault detection scenarios, models must operate on data in real time with little latency. While KNN has greater inference time due to the computation required in distances, the 'Fine' mode, which limits the number of neighbors and operates in a lowerdimensional feature space (after feature selection), significantly reduces the computational overhead. Moreover, unlike ensemble methods, which involve combining multiple learners and can cause inference latency, Fine KNN involves a simple and direct classification process.[18]

Quadratic SVM, although precise, involves computationally costly kernel computations for large data sets or high-dimensional data. Bagged ensembles, although robust, aggregate predictions from several base learners, further increasing inference time. Fine KNN thus balances accuracy and computational simplicity, thus more real-time deployable on devices with limited processing power.

5. Discussion

The Fine KNN machine learning model consisted of 85,380 feature values with 7,115 observations as each contained 13 feature values. Originally, the used dataset contained 16 features; however, it was noted that the skewness, margin factor and mean are less relevant features for further analysis and, therefore, excluded from the study. The features mean, median, skewness, and margin factor ranked lowest in terms of importance in the feature ranking with the Relief algorithm. Upon deletion of the four features and retraining of the model, it achieved 95.6% accuracy. Although the median was identified as a borderline feature, removal of it, helped to improve the accuracy. The accuracy of the model improved from 94.1% in the development phase to 95.6 % which is a notable improvement after the feature reduction and can be seen in Figure 15. This improvement in accuracy indicates that the reduced feature set provided a clearer and more relevant signal for classification.

The confusion matrix highlights that most misclassifications were related to bearing defects, suggesting that this fault type was less accurately learned by the model compared to other conditions such as healthy states, gear faults, bearing faults, or fan faults.

The proposed model obtained an accuracy, of approximately 95.6% percent, which is relatively similar to the one obtained by Altaf et al. [2], yet, their work was limited to bearing faults only. However, this research integrates a more diverse range of faults such as bearing, fan, and gear indicating that the model is not restricted to certain conditions only.

A comparison of the findings of this study with those from previous research by Shubita, Alsadeh, and Khater [4] That used similar research methods such as Acoustic Emission techniques, Machine Learning, and IoT-based real-time monitoring of faults revealed disparities. Although previous studies have shown very high accuracy in fault classification of 96.1%, the current study yielded different results. The variations between this research and past work show the problem of deploying machine learning techniques to real-world industrial domains and the external factors that can influence the capabilities of the diagnostic system. It is clear that these deviations highlight some problematic aspects in designing the fault detection systems and can show which aspects need to be further investigated to obtain better performance of AE and ML integration under various circumstances.

In comparison with the research undertaken by Kiran Vernekar and KV Gangadharan [11] They have got accuracy of about 85.5% in detecting gear and bearing faults in a gearbox system. However, their approach did not take into account edge learning in machines and did not consider the effects of vibrations caused by the combustion engine.

On the other hand, (Swapnil K. Gundewar and Prasad V. Kane[6] Research showing the highest accuracy for bearing fault diagnosis methods, but tied to high demand on computations makes it impractical for real-time applications in resource-constrained embedded systems.

Capabilities in automatic fault detection have reached new heights owing to machine learning methods. Cited in the [19], ML algorithms like SVM, ANN, CNN, and RNN have their strong points when applied in predictive maintenance. Likewise, Kumar et al. [20] Reported employing SVM, ANN, and Bayesian classifiers for fault detection of drilling machines, emphasizing feature extraction and selection to boost performance.

To that effect Nguyen et al. [21] Assessed ML models for diagnosing unbalanced and misalignment faults and concluded that Random Forest was the most applicable. In the same way, Jolfaei et al. [22] Made use of the random forest algorithms for high-voltage water pumps with a diagnostic accuracy of 97% in real-time. Mayaki and Riveill [23] Suggested the use of feature extraction techniques to make deep neural networks suitable for 2D images. In their work, they demonstrated that with minimal data, 100% accuracy could be attained in the early stages of fault detection.

The challenges of fault detection for the complex deep learning approaches have progressed as well. Panigrahi et al. [24] Focused on the performance of CNNs and RNNs for a variety of industrial fault cases remarking that such neural networks are robust against sensor misbehaviors and noise. Yu et al. [25] Suggested the use of one-dimensional convolutional neural networks (1D-CNN) for vibration signals and obtained a high accuracy as well as low costs and reliance on hand-engineered features.

As Hongmei Liu, Lianfeng Li and Jian Ma [3] Observe that this flexibility has the added consequence that deep learning models do not necessarily have to go through the process of feature selection only to achieve reasonable accuracy. However, it is generally understood that depth learning methods entail higher memory and fluctuating performances over the operational conditions. This research however had to adopt the RelieF algorithm for feature ranking, as this reduced the number of features and enhanced the classifier outcomes. Due to the minimum subset of features, stable accuracy and real-time prediction capability were obtained, which is suitable for the edge ML model.

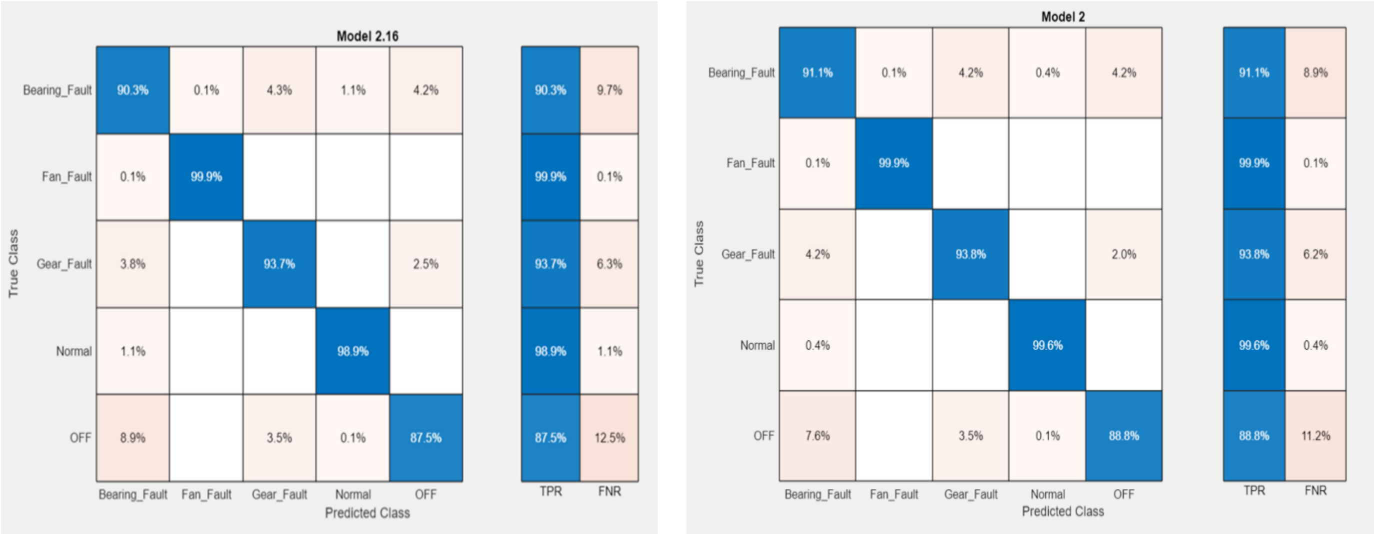


Fig.18. Comparison of confusion matrices for the Fine KNN model before and after feature reduction.

The proposed Fine KNN model does quite well, giving an accuracy of 95.6 % in fault detection and diagnosis. This method provides a strong possibility for classifying various faults in the rotating equipment, therefore making a great potential for practical use. However, it is future work that could compare the use of more elaborate classifiers like the Bagged classifiers or SVMs Quadratic since they are seen in preliminary research to be more precise and more resistant. The latter models may give even higher accuracy and robustness in fault diagnosis tasks.

Moreover, from the findings of the research study, Fine KNN is equally shown to serve the purpose of selecting the most relevant features besides offering high performance. To enrich the model, future studies can consider the experiments based on the different preprocessing approaches, such as those mentioned by Qiao and Shu. [26]. It can be seen these methods aim at learning useful information from the noisy signals and they can pave the way to increase the model accuracy and also the stability of the model. Future developments can also adapt such sophisticated methods to enhance the detection and diagnosis functions of the model.

This study utilized openly accessible data to train the machine learning models rather than original data being gathered in a laboratory setting. The reason for not gathering original data was that machinery with various types of faults was not available. It was just a binary difference—whether a fault exists or not—that could have been experimented on under available conditions. However, the primary objective of this research is to perform multi-class fault classification and not merely fault detection. Testing for the existence of a fault alone would be beyond the scope of this project.

The importance of conducting experiments to obtain original data in the laboratory is recognized. Such validation would introduce realistic complexities and variations, resulting in a more comprehensive model evaluation. In the future, we will overcome these limitations by creating an experimental setup that is capable of simulating and capturing a range of fault types, thus enabling full experimental validation of the proposed models.

This research paves a good foundation for future integration into real-time edge devices as a precursor to fostering practical industrial usage. In addition to showing how Fine KNN works and performs, the methodologies and findings contribute not only to the considerations of better classifiers and preprocessing but also to the future development of proposing more advanced methodologies. This evolution will be instrumental in the formulation of better approaches to Fault detection systems in the industrial processes in terms of improvements: sophistication, reliability, and real-time monitoring.

6. Conclusions

This paper presents the development of a fault detection and diagnosis model for three rotating elements of a commercial drill tool the bearing, fan, and gear. The process progressed from idea to an early implementation stage, with numerous machine learning methods and architectures tested to find the best classifier. It was concluded that the fine KNN model has the finest trade-off between accuracy and computational time and therefore was selected for the particular application.

The entire diagnostic process, from data acquisition and preprocessing to the extraction of features and the classification of faults, was performed during the development phase. The fine KNN model which is chosen demonstrated high accuracy and stability, so it can be stated that the chosen fine KNN model can become a perfect solution for the diagnosis of machine faults based on the sound signals.

To ensure that a fault diagnosis system that uses machine learning is implemented properly the following factors must be observed. The first condition is that there must be a sufficient amount of observational data related to the identified fault conditions, as it improves the model’s training dataset and generalization capabilities when considering various operating modes. Second, there is the need to ensure that high levels of preprocessing algorithms result in better-quality data. Methods like windowing with overlap and digital low/high pass filtering aid in attenuation of signal noise which is crucial for feature extraction.

Third, it is important to choose the right feature extraction techniques and apply them to achieve high diagnostic accuracy. In this study, the fine KNN model was used because it provides high accuracy and is not resource-demanding. The high accuracy and stability established in The model serves to support its application to real-world problems. The ability to diagnose faults from sound signals, coupled with operational stability adds credibility to the model for online machine condition monitoring. Besides, this approach paves the way to develop a device that can also prevent faults in real-time, and it is beneficial for constructing an economical and efficient monitoring system, which enhances the applicability of this method in industrial environments.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] L. C. Brito, G. A. Susto, J. N. Brito, and M. A. V. Duarte, 'An explainable artificial intelligence approach for unsupervised fault detection and diagnosis in rotating machinery', *Mech Syst Signal Process*, vol. 163, p. 108105, Jan. 2022, doi: 10.1016/j.ymssp.2021.108105.

[2] M. Altaf *et al.*, 'Automatic and Efficient Fault Detection in Rotating Machinery using Sound Signals', *Acoust Aust*, vol. 47, no. 2, pp. 125–139, Aug. 2019, doi: 10.1007/s40857-019-00153-6.

[3] Hongmei Liu, Lianfeng Li, and Jian Ma, 'Rolling bearing fault diagnosis based on stft-deep learning and sound signals', *Shock and Vibration*, pp. 345–357, 2016.

[4] R. R. Shubita, A. S. Alsadeh, and I. M. Khater, 'Fault Detection in Rotating Machinery Based on Sound Signal Using Edge Machine Learning', *IEEE Access*, vol. 11, pp. 6665–6672, 2023, doi: 10.1109/ACCESS.2023.3237074.

[5] A. Senanayaka, P. Lee, N. Lee, C. Dickerson, A. Netchaev, and S. Mun, 'Enhancing the Accuracy of Machinery Fault Diagnosis through Fault Source Isolation of Complex Mixture of Industrial Sound Signals', May 13, 2024. doi: 10.21203/rs.3.rs-4365283/v1.

[6] Swapnil K. Gundewar and Prasad V. Kane., 'Rolling element bearing fault diagnosis using supervised learning methods- artificial neural network and discriminant classifier.', *International Journal of System Assurance Engineering and Management*, 2022.

[7] A. F. Khalil and S. Rostam, 'Machine Learning-based Predictive Maintenance for Fault Detection in Rotating Machinery: A Case Study', *Engineering, Technology & Applied Science Research*, vol. 14, no. 2, pp. 13181–13189, Apr. 2024, doi: 10.48084/etasr.6813.

[8] I. M. Brandao and C. da Costa, 'Fault Diagnosis of Rotary Machines Using Machine Learning', *Eletrônica de Potência*, vol. 27, no. 3, pp. 236–243, Aug. 2022, doi: 10.18618/rep.2022.3.0013.

[9] T. Tran, N. T. Pham, and J. Lundgren, 'Detecting Drill Failure in the Small Short-sound Drill Dataset', Aug. 2021.

[10] C. Vununu, K.-S. Moon, S.-H. Lee, and K.-R. Kwon, 'A Deep Feature Learning Method for Drill Bits Monitoring Using the Spectral Analysis of the Acoustic Signals', *Sensors*, vol. 18, no. 8, p. 2634, Aug. 2018, doi: 10.3390/s18082634.

[11] Kiran Vernekar and KV Gangadharan, 'Engine gearbox fault diagnosis using machine learning approach', *J Qual Maint Eng*, vol. 24, no. 3, 2018.

[12] S. Liu, M. Gao, Y. Feng, and L. Sheng, 'Dynamic event-triggered fault detection for rotary steerable systems with unknown time-varying noise covariances', *ISA Trans*, vol. 142, pp. 478–491, Nov. 2023, doi: 10.1016/j.isatra.2023.08.018.

[13] Jonguk Lee, Heesu Choi, Yongwha Chung, Hee-Young Kim, and Sukhan Yoon, 'Fault Detection and Diagnosis of Railway Point Machines by Sound Analysis', *Sensors*, vol. 16, no. 4, 2016.

[14] 'RashadShubita/Fault-Detection-using-TinyML: Fault Detection in Rotating Machinery based on Sound Signal using Edge Machine Learning'. Accessed: Apr. 23, 2025. [Online]. Available: https://github.com/RashadShubita/Fault-Detection-using-TinyML

[15] 'Audio-Based Anomaly Detection for Machine Health Monitoring'. Accessed: Apr. 23, 2025. [Online]. Available: https://in.mathworks.com/help/audio/ug/audio-based-anomaly-detection-for-machine-health-monitoring.html

[16] 'Feature Extraction Explained - MATLAB & Simulink'. Accessed: Apr. 23, 2025. [Online]. Available: https://in.mathworks.com/discovery/feature-extraction.html

[17] 'SVM vs KNN in Machine Learning | GeeksforGeeks'. Accessed: Apr. 22, 2025. [Online]. Available: https://www.geeksforgeeks.org/svm-vs-knn-in-machine-learning/

[18] 'ML | Bagging classifier | GeeksforGeeks'. Accessed: Apr. 22, 2025. [Online]. Available: https://www.geeksforgeeks.org/ml-bagging-classifier/

[19] G. Ciaburro, 'Machine fault detection methods based on machine learning algorithms: A review', *Mathematical Biosciences and Engineering*, vol. 19, no. 11, pp. 11453–11490, 2022, doi: 10.3934/mbe.2022534.

[20] A. Kumar, J. Ramkumar, N. K. Verma, and S. Dixit, 'Detection and classification for faults in

drilling process using vibration analysis', in *2014 International Conference on Prognostics and Health Management*, IEEE, Jun. 2014, pp. 1–6. doi: 10.1109/ICPHM.2014.7036393.

[21] N. H. T. Nguyen, P. N. H. Nguyen, T. M. Le, Q. N. Truong, L. TonThat, and S. V. T. Dao, 'Fault Diagnosis for Rotating Machine Utilizing Machine Learning with FFT-based Feature Extraction', in *2024 Tenth International Conference on Communications and Electronics (ICCE)*, IEEE, Jul. 2024, pp. 619–624. doi: 10.1109/ICCE62051.2024.10634660.

[22] N. G. Jolfaei *et al.*, 'An Intelligent Dynamic Fault Diagnostic Model for Rotating Machinery', in *2023 7th International Conference on System Reliability and Safety (ICSRS)*, IEEE, Nov. 2023, pp. 407–412. doi: 10.1109/ICSRS59833.2023.10381093.

[23] M. Z. A. Mayaki and M. Riveill, 'Machinery Anomaly Detection using artificial neural networks and signature feature extraction', in *2023 International Joint Conference on Neural Networks (IJCNN)*, IEEE, Jun. 2023, pp. 1–10. doi: 10.1109/IJCNN54540.2023.10191814.

[24] B. S. Panigrahi, T. T, M. Tamilselvi, S. B. G. Tilak Babu, P. G, and B. Shaik, 'Deep Learning Techniques for Fault Detection in Industrial Machinery', in *2024 5th International Conference on Recent Trends in Computer Science and Technology (ICRTCST)*, IEEE, Apr. 2024, pp. 221–226. doi: 10.1109/ICRTCST61793.2024.10578499.

[25] F. Yu, G. Chen, C. Du, L. Liu, X. Xing, and X. Yang, 'Fault diagnosis and identification of rotating machinery based on one-dimensional convolutional neural network', *Journal of Vibroengineering*, vol. 26, no. 4, pp. 793–807, Jun. 2024, doi: 10.21595/jve.2024.23722.

[26] Z. Qiao and X. Shu, 'Coupled neurons with multi-objective optimization benefit incipient fault identification of machinery', *Chaos Solitons Fractals*, vol. 145, p. 110813, Apr. 2021, doi: 10.1016/j.chaos.2021.110813.