PREDICTIVE MAINTENANCE IN MACHINE TOOLS

PROJECT REPORT

18CSE483T – Intelligent Machining

(2018 Regulation)

III Year/ VI Semester

Academic Year: 2023 -2024

By

SARTHAK DWIVEDI (RA2111026010122)

SHASHWAT PRASAD (RA2111026010139)

YASHOWARDHAN SAMDHANI (RA2111026010151)

Under the guidance of

Dr. A. ROBERT SINGH

Assistant Professor

Department of Computational Intelligence

FACULTY OF ENGINEERING AND TECHNOLOGY

SCHOOL OF COMPUTING

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

Kattankulathur, 603203

MAY 2024

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

(Under Section 3 of UGC Act, 1956)

BONAFIDE CERTIFICATE

Certified that Mini project report titled "**PREDICTIVE MAINTENANACE IN MACHINE TOOLS**" is the bona fide work of **SARTHAK DWIVEDI (RA2111026010122), SHASHWAT PRASAD (RA2111026010143), YASHOWARDHAN SAMDHANI (RA2111026010151)** who carried out the minor project under my supervision. Certified further, that to the best of my knowledge, the work reported herein does not form any other project report or dissertation based on which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE

Dr. A. Robert Singh

IM – Course Faculty

Assistant Professor

Department of Computational Intelligence

SRM Institute of Science and Technology

Kattankulathur

SIGNATURE

Dr. R. Annie Uthra

Head of the Department

Professor

Department of Computational Intelligence

SRM Institute of Science and Technology

Kattankulathur

ACKNOWLEDGEMENT

We express our humble gratitude to Dr. C. Muthamizhchelvan, Vice-President Chancellor, SRM Institute of Science and Technology, for the facilities extended for the project work and his continued support. We extend our sincere thanks to Dean-CET, SRM Institute of Science and Technology, Dr. T. V. Gopal, for his invaluable support.

We wish to thank Dr. Revathi Venkataraman, Professor & Chairperson, School of Computing, SRM Institute of Science and Technology, for her support thorough out the project work. We are incredibly grateful to our Head of the Department, Dr. R. Annie Uthra Professor, Department of Computational Intelligence, SRM Institute of Science and Technology, for her suggestions and encouragement at all the stages of the project work.

We wish to thank Dr. A. Robert Singh, Assistant Professor Department of Computational Intelligence, School of Computing, SRM Institute of Science and Technology, for his support throughout the project work.

We register our immeasurable thanks to our Faculty Advisor, Dr. A. Saranya, Assistant Professor, Department of Computational Intelligence, School of Computing, SRM Institute of Science and Technology, for leading and helping us to complete our course.

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	LIST OF FIGURES	V
	ABSTRACT	VI
1	INTRODUCTION	7-8
2	LITERATURE SURVEY	9-11
3	PROBLEM STATEMENT	12
4	INDUSTRIAL MAINTENANCE & MACHINE TOOLS	13-17
5	RELEVANT WORKS	18-20
6	DATASET AND FAULTS	21-23
7	WORKING PRINCIPLE	24-27
8	RESULT	28-36
9	DISCUSSION	37-40
10	CONCLUSION	41
	FUTURE SCOPE	42-43
	REFERENCES	44-45

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE NO.
4.1	TYPES OF INDUSTRIAL MAINTENANCE	13
4.2	REACTIVE MAINTENANCE OVERVIEW	14
4.3	PREVENTIVE MAINTENANCE OVERVIEW	15
4.4	PREDICTIVE MAINTENANCE OVERVIEW	16
4.5	SUPERVISED LEARNING	17
4.6	UNSUPERVISED LEARNING	17
4.7	REINFORCEMENT LEARNING	17
5.1	VIBRATION TRAVEL DATA	18
5.2	OIL ANALYSIS DATA	19
5.3	INFRARED DATA OF MOTOR STATER	20
6.1	ROTATORY MACHINERY DATA PERCENTAGE IN MACHINE FAULT DATABASE	22
6.2	SUMMARY OF 6 STATES OF ROTATORY MACHINES MEASUREMENT	22
7.1	MACHINE LEARNING AND DEEP NEURAL NETWORK PIPELINE FOR GEARBOX AND ROTATORY MACHINING	25
7.2	GEARBOX DATA ACQUISITION	25
7.3	ROTATORY MACHINING DATA ACQUISITION	26
7.4	PREPROCESSING PIPELINE	26
7.5	DNN ARCHITECTURE FOR GEARBOX	27
8.1	RATIO OF TRAINING AND TEST DATA	28
8.2	DATA DISTRIBUTION AMONG NORMAL AND BROKEN GEARBOX CLASSES	28
8.3	PERFORMANCES EVALUATION OF ML, DNN MODEL ON RAW DATA	29
8.4	ROC CURVE AND CORRESPONDING AUC SCORE OF BOTH ML AND DNN MODEL ON RAW DATA	30
8.5	PERFORMANCE EVALUATION OF ML AND DLL MODEL WITH N=10	31
8.6	PERFORMANCE EVALUATION OF ML AND DNN MODELS WITH N=100	32
8.7	ROC CURVE AND CORRESPONDING AUC SCORE OF BOTH ML AND DNN MODELS WITH N=100	33
8.8	THE NUMBER OF RECORDS IN EACH OF SIX CLASSES IN MFP	34
8.9	PERFORMANCE EVALUATION OF RF OF MFP USING A CONFUSION MATRIX	35
8.10	ROC CURVE AND CORRESPONDING AUC SCORE OF RF MODEL IN MFP DATASET	36

ABSTRACT

This study delves into the application of advanced machine learning techniques to pre-empt losses and faults in industrial machinery, thereby paving the way for predictive maintenance. Utilizing two distinct datasets focusing on gearbox and rotary machinery, the research aims to predict faults employing machine learning and deep neural network models. Evaluation of model performance encompasses binary and multi-classification problems, employing various machine learning algorithms and assessing their statistical metrics. Results indicate that for the gearbox fault dataset, both random forest and deep neural network models exhibited comparable performance, achieving high F1-scores, AUC scores, and minimal error rates. Conversely, in the multi-classification rotatory machinery fault prediction dataset, the random forest model outperformed the deep neural network model. In summary, the study underscores the efficacy of machine learning and deep neural network models, particularly random forest and deep neural network, in accurately predicting various types of rotatory machinery and gearbox faults compared to other algorithms such as decision tree and AdaBoost.

Keywords:

Machine Learning, Deep Learning, Big Data, Predictive Maintenance, Rotatory Machinery Fault Prediction, Gearbox Fault Prediction, Machinery Fault Database, Internet of Things (IoT), Spectra quest machinery fault simulator, Cloud Computing, Industry 4.0

CHAPTER – 1

INTRODUCTION

The field of industrial maintenance has undergone significant transformation in recent years, driven by the integration of advanced analytics techniques and machine learning algorithms. The motivation behind such advancements stems from the pressing need to address the challenges posed by maintaining various types of industrial machinery efficiently and effectively.

In today's world, where machines are indispensable across numerous sectors, including automotive, manufacturing, oil and gas, renewable energy, and more, ensuring their smooth operation is paramount. However, the complexity of modern machinery, coupled with the high stakes involved in their maintenance, presents formidable challenges. Not only does unplanned downtime due to machinery failures incur substantial costs in terms of repairs and lost productivity, but it can also jeopardize entire production lines and plant operations.

Against the backdrop of the COVID-19 pandemic, industries have accelerated their digital transformation efforts, seeking to optimize operations and minimize manual intervention. This shift towards digitization underscores the importance of leveraging advanced analytics techniques to enable predictive maintenance strategies. By harnessing technologies such as cloud computing, big data analytics, artificial intelligence, machine learning, and the Internet of Things (IoT), industries can proactively monitor the health of their assets and pre-emptively address potential issues before they escalate.

Predictive maintenance, once primarily associated with the oil and gas industry, has now permeated various domains, thanks to the advent of IoT and cutting-edge technologies. By harnessing data generated by sensors embedded within machinery, predictive maintenance algorithms can forecast equipment failures, allowing maintenance teams to intervene proactively. Automotive industries are transitioning from reactive maintenance approaches to predictive models, recognizing the immense value in pre-emptively addressing potential faults.

The predictive maintenance framework outlined in the provided case study addresses these challenges comprehensively. By employing advanced analytics techniques, such as vibration analysis, oil analysis, motor current analysis, and infrared thermography, maintenance teams can glean valuable insights into the health and performance of industrial machinery. These techniques enable the early detection of potential faults, ranging from gear meshing defects and bearing failures to imbalance conditions and insulation breakdowns in electric motors.

Furthermore, the integration of various predictive maintenance technologies facilitates a holistic approach to asset management, allowing maintenance teams to address multiple failure modes simultaneously. By identifying root causes and implementing corrective measures, industries can enhance the reliability, efficiency, and longevity of their equipment, thereby minimizing downtime and maximizing productivity.

In essence, the convergence of advanced analytics, machine learning, and IoT has ushered in a new era of predictive maintenance, empowering industries to proactively manage their assets and optimize operational performance. By leveraging data-driven insights and predictive algorithms, maintenance teams can transform reactive maintenance practices into proactive strategies, ensuring the seamless operation of industrial machinery and safeguarding business continuity in an increasingly competitive landscape.

CHAPTER - 2
LITERATURE SURVEY

A Literature survey on Predictive Maintenance in Machine Tools

Predictive maintenance (PdM) methodologies have become indispensable in modern industrial settings, offering proactive strategies for equipment health management. This survey delves into the various approaches within PdM, including vibration analysis, oil analysis, infrared thermography, and motor current analysis, highlighting their advantages and disadvantages.

Vibration Analysis

Vibration analysis stands as a cornerstone of PdM, offering insights into machinery health by detecting gear meshing defects, bearing faults, and imbalance conditions. Furthermore, vibration analysis extends to resonance detection, crucial for optimizing operating conditions and avoiding catastrophic failures due to natural frequency excitation.

Advantages:

- Early detection of gear meshing defects, bearing faults, and imbalance conditions.
- Enables pre-emptive corrective actions, minimizing costly downtime.
- Facilitates resonance detection, optimizing operating conditions and preventing catastrophic failures.

Disadvantages:

- Complexity in diagnosing faults, particularly in intricate gear configurations.
- Requires specialized training and expertise for accurate interpretation of data.
- Limited effectiveness in detecting non-mechanical issues.

Oil Analysis

Oil analysis emerges as another vital diagnostic tool within PdM. By assessing oil cleanliness and identifying contaminants, such as wear metals, this methodology facilitates early detection of component degradation. Maintaining optimal oil conditions through filtration and leak prevention becomes imperative for prolonging equipment lifespan and minimizing wear-related failures.

Advantages:

- Early detection of contaminants and wear metals, facilitating proactive maintenance.
- Prolongs equipment lifespan by maintaining optimal oil conditions through filtration.
- Identifies bearing and gear problems that could otherwise go undetected.

Disadvantages:

- Relies on regular sampling and analysis, which can be time-consuming and costly.
- Limited ability to diagnose electrical or non-mechanical issues.
- Requires careful interpretation to distinguish between normal wear and abnormal conditions.

Infrared Thermography

Infrared thermography presents a non-invasive means of assessing equipment health by capturing temperature differentials indicative of potential issues. With applications ranging from motor control starters to high-voltage power lines, infrared thermography aids in detecting anomalies like loose connections, which, if left unaddressed, could lead to premature equipment failure.

Advantages:

- Non-invasive assessment of equipment health through temperature differentials.
- Detects anomalies such as loose connections, facilitating preventive maintenance.
- Applicable across various equipment types, from low-voltage to high-voltage systems.

Disadvantages:

- Dependent on ambient temperature and environmental conditions.
- Limited effectiveness in diagnosing internal faults or mechanical wear.
- Requires trained personnel for accurate data interpretation and analysis.

Motor Current Analysis

Motor current analysis complements vibration and infrared analyses, providing insights into electrical and mechanical health. By monitoring parameters like resistance imbalance and insulation breakdown, analysts can predict and prevent failures in electric motors, mitigating the risk of unscheduled downtime and costly repairs.

Advantages:

- Predicts insulation breakdown and mechanical failures in electric motors.
- Complements vibration and infrared analyses, providing comprehensive health insights.
- Offers insights into air gap anomalies and rotor-stator rubs.

Disadvantages:

- Limited to electric motor diagnostics, excluding other equipment types.
- Requires specialized equipment and expertise for accurate measurements.
- May necessitate motor disassembly for detailed analysis, leading to downtime.

CHAPTER – 3
PROBLEM STATEMENT

To develop an advanced predictive maintenance system tailored for industrial equipment, integrating multiple sensing technologies to detect and predict various failure modes proactively. By incorporating vibration analysis, oil analysis, infrared thermography, and motor current analysis, the system aims to provide comprehensive monitoring and diagnostics capabilities for detecting gear meshing defects, bearing failures, imbalance conditions, loose connections, and insulation breakdown in electric motors. The primary goal is to enable timely maintenance interventions to prevent catastrophic failures, minimize downtime, and extend the lifespan of critical machinery. Through the implementation of a robust control algorithm, the system will process real-time sensor data, make intelligent decisions, and trigger maintenance actions based on predictive analytics, ensuring optimal equipment performance and reliability. Additionally, the project includes the development of a user-friendly interface for monitoring equipment health, visualizing sensor data, and generating maintenance reports, facilitating easy deployment and utilization by maintenance personnel. Comprehensive testing and validation in both simulated and real-world industrial environments will be conducted to assess the system's reliability, accuracy, and effectiveness. Ultimately, the proposed predictive maintenance system aims to revolutionize maintenance practices in industrial settings, ushering in a new era of proactive equipment health management and operational efficiency.

CHAPTER – 4

INDUSTRIAL MAINTENANCE AND MACHINE TOOLS

4.1. Maintenance

The maintenance cost in many industries is higher than operational and production costs due to premature equipment failure [9]. The profitability of any industry generally depends on the maintenance process. Normally maintenance in industries happens when the equipment reaches a certain age or stops working [10]. It is good to do scheduled maintenance, but it does not provide any information about the equipment's health in the future. To optimize the production lines and equipment reliability, different types of maintenance can be performed based on the resource. The most common types of industrial maintenance are Figure 4.1

1. Reactive Maintenance
2. Preventive Maintenance
3. Predictive Maintenance

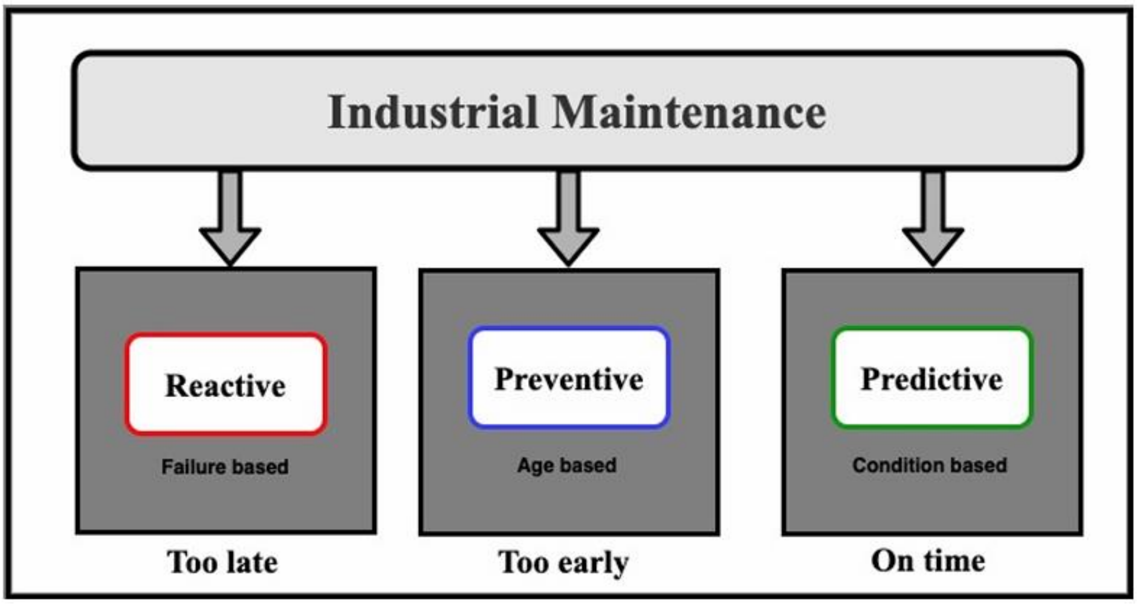

Figure 4.1: Types of Industrial Maintenance

4.1.1. Reactive Maintenance

In this approach, maintenance can be performed when components or machinery have a problem or stop working. Normally maintenance will perform after the equipment failure as shown in Figure 4.2. Although the component or machine is used full lifespan, drawbacks of this approach are

- Unscheduled maintenance
- Downtime is increased

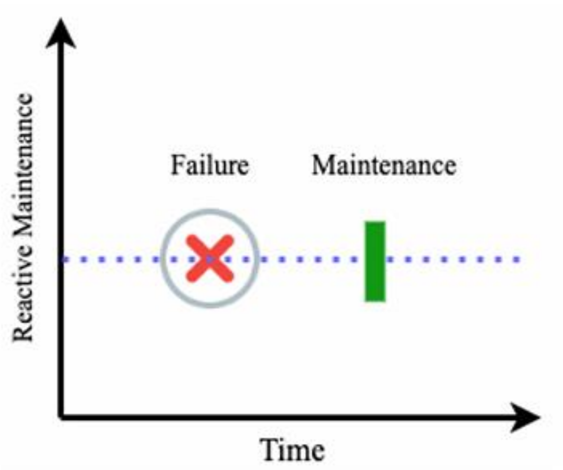

Figure 4.2: Reactive Maintenance Overview

4.1.2. Preventive Maintenance

In this approach, the machine or component is replaced in advance before it fails. It helps to avoid unscheduled maintenance. The maintenance will perform during the regular interval as shown in Figure 4.3. The drawback of this approach is [11,12,13]

- The component or machine is not fully utilized
- Over maintenance is performed

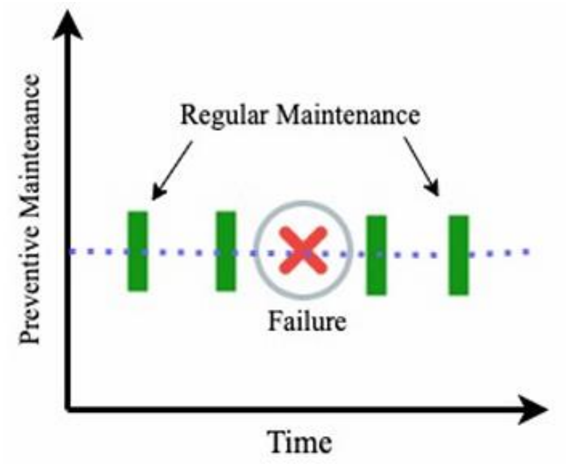

Figure 4.3: Preventive Maintenance Overview

The drawbacks of regular maintenance are

- Breakdown time is increased
- Productivity is reduced due to regular maintenance
- Over maintenance of some equipment or machinery
- Operation cost is an increase
- More skilled labour is needed to maintain the equipment

4.1.3. Predictive Maintenance

It predicts the fault and performs the maintenance on the machine or equipment before the fault or failure happens as shown in the Figure 4.4. It extends the life span of the equipment. There are several advantages of predictive maintenance [13,14,15] such as,

- It can reduce the unplanned downtime
- It can help to identify fault or equipment health by condition monitoring to avoid costly equipment failure
- It decreased the planned downtime by reducing inspection and premature repair

Predictive maintenance system is an IoT based system. The drawback of this approach is the initial cost to build such a system is very high.

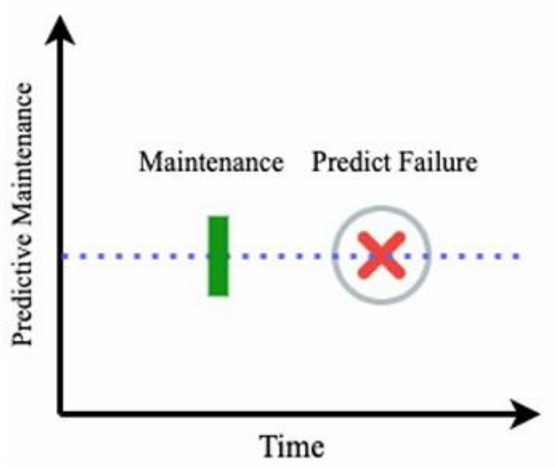

Figure 4.4: Predictive Maintenance Overview

4.2. Machine Learning (ML)

IoT and cloud computing make machine learning possible in manufacturing and other industries. Now it is much easier to get the data from the industrial equipment with IoT devices. ML transforms some of the tasks to a machine that was previously not possible with humans [16].

4.2.1 Types of Machine Learning

The ML is of three types

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning (RL)

1. Supervised Learning

Supervised learning techniques are easy to understand and implement. Labelled data is provided to the ML models [17,18]. It means both training and validation data are labelled. The training datasets comprise both inputs and target outputs in supervised learning as shown in Figure 4.5. It can be used for both classification and regression problems. The algorithms in supervised learning are decision trees, random forest, support vector machine, navies byes, linear regression, logistic regression, etc.

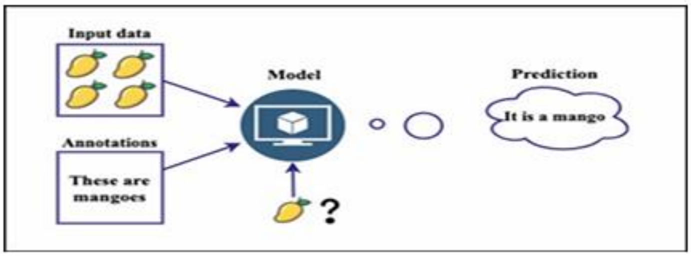

Figure 4.5: Supervised Learning

2. Unsupervised Learning

In this approach the user does not need to provide the label data to the model, it works with unlabelled data [19]. It allows the model to detect patterns and information on its own Figure 4.6. The algorithms in unsupervised learning are clustering, K-Nearest Neighbors (KNN), anomaly detection, Principal Component Analysis (PCA), etc.

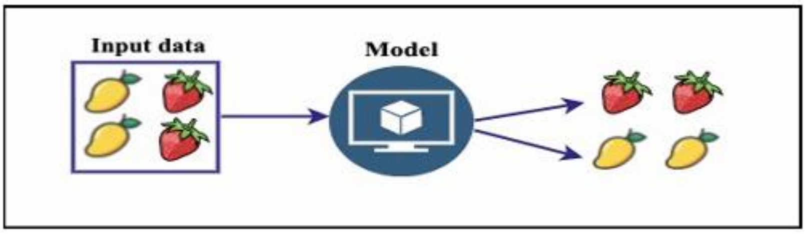

Figure 4.6: Unsupervised Learning

3. Reinforcement Learning

RL is a type of ML and does not require a lot of training data. Instead of environments are given to the RL models, the agent learns from its environment by trial and error to achieve goals and get rewards Figure 4.7.

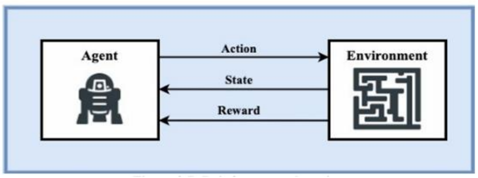

Figure 4.7: Reinforcement Learning

CHAPTER – 5
RELEVANT WORKS

The realm of predictive maintenance by presenting a detailed exploration of various case studies and the application of different technologies in industrial settings. It begins by discussing the importance of predictive maintenance in preventing equipment failures and increasing uptime, setting the stage for the subsequent case studies. Each case study is meticulously crafted to demonstrate the effectiveness of specific predictive maintenance techniques, such as vibration analysis, oil analysis, infrared thermography, and motor current analysis.

The vibration analysis case study, for instance, provides a comprehensive examination of gear meshing defects, bearing faults, and imbalance conditions using spectral and time waveform analysis. It elucidates how the identification of subtle abnormalities in vibration patterns can enable early detection of potential failures, thereby averting catastrophic consequences and reducing repair costs.

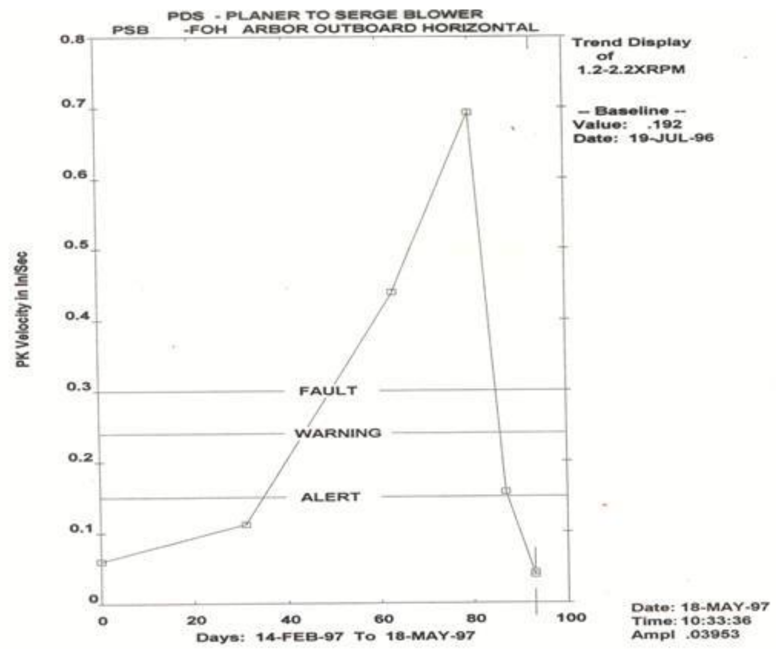

Figure 5.1: Vibration Trend Data

Similarly, the oil analysis case study underscores the critical role of oil cleanliness in preventing premature component failure. By analysing oil samples for contaminants, maintenance personnel can assess equipment health and identify underlying issues such as wear debris or ingresses dirt, thus extending the lifespan of machinery and enhancing operational reliability.

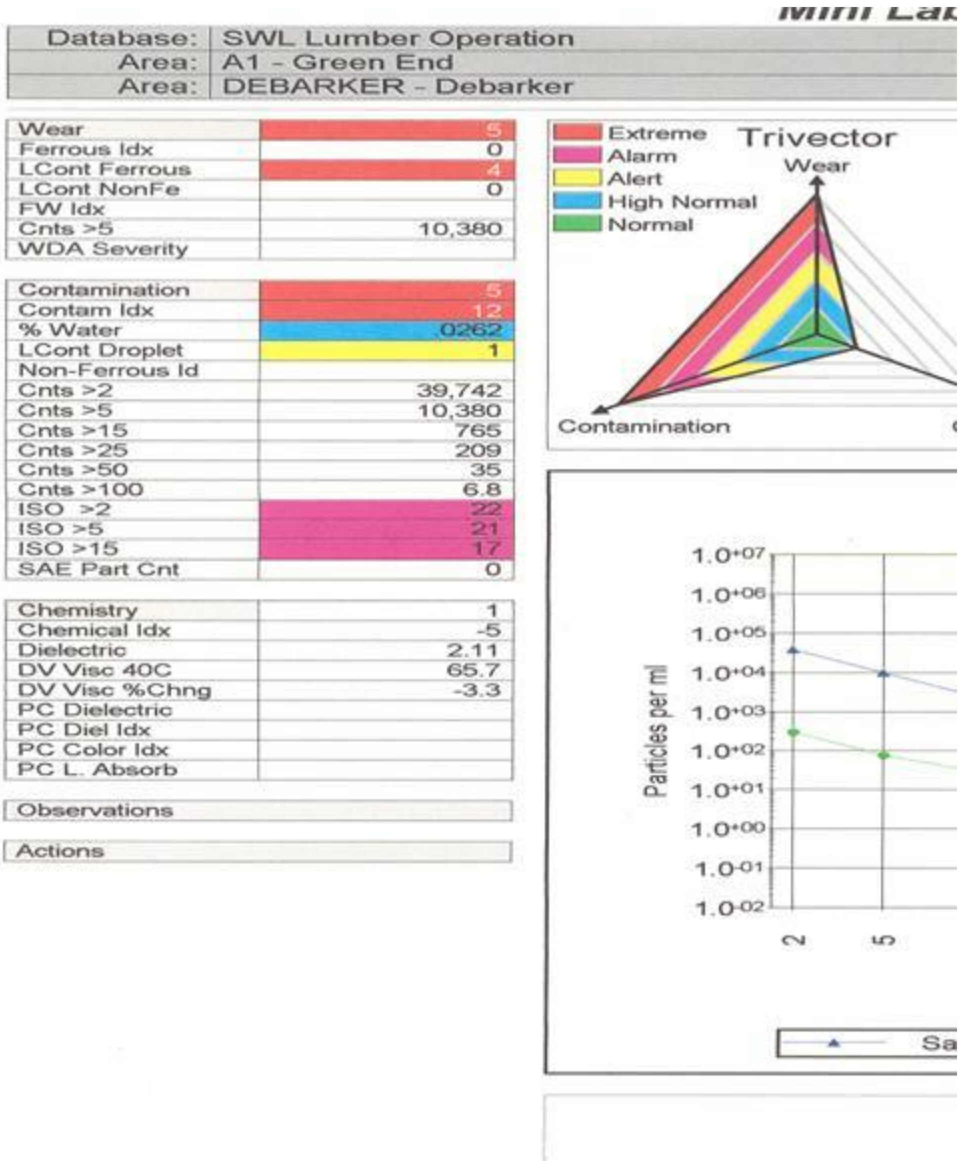

Figure 5.2: Oil Analysis Data

The infrared thermography case study showcases the ability of thermal imaging to detect temperature variations indicative of underlying problems, such as loose connections or overheating components. By capturing thermal images and interpreting temperature data, technicians can pinpoint potential issues and take proactive measures to address them before they escalate into major failures.

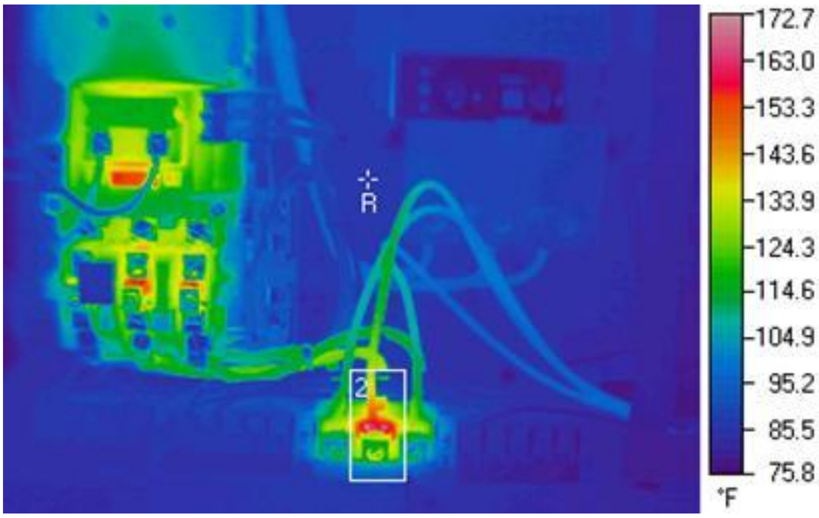

Figure 5.3: Infrared Data of Motor Starter

Furthermore, the motor current analysis case study delves into the intricacies of diagnosing insulation breakdown and detecting air gap anomalies in electric motors. Through meticulous data analysis and comparison, maintenance professionals can identify early signs of motor degradation and implement timely interventions to prevent costly downtime.

Overall, the holistic approach to predictive maintenance, advocating for the integration of multiple technologies and the importance of skilled personnel in interpreting data and implementing preventive measures effectively. By showcasing real-world examples and highlighting the tangible benefits of predictive maintenance, the article underscores its significance in modern industrial operations and maintenance practices.

CHAPTER – 6

DATASET AND FAULTS

Data is the core component of any ML/DL model. Quality data is required to perform these models efficiently. The performance of the ML/DL model can improve by integrating more data into the ML/DL system. The data can be of many forms, but the ML model mainly rely on

- Numerical data
- Text data
- Categorical data
- Time series data

6.1 Experimental Setup

Spectra quest provides different types of simulators for training and studying industrial machine behaviours. These simulators accelerate learning and help to understand the different types of faults in industrial machinery [20]. The data we used to train and test the ML model was taken from these simulators

- Spectra Quest's Gearbox Fault Diagnostics Simulator
- Spectra Quest's Machinery Fault Simulator

6.2 Gearbox Dataset

The gearbox dataset used in this study is publicly available at Open Ei [21]. The data was recorded by Open Ei [21] with the four vibration sensors placed in different directions on spectra quests gearbox fault diagnostics simulator [20]. The dataset is recorded with a different load from 0 to 90 percent and contains information about the health conditions of the gearbox based on the vibrational sensors reading. Gearbox dataset describes only two states of gearbox such as

- Normal
- Broken teeth

6.3 Machinery Fault database

The data from spectra Quest Machinery Fault Simulator (MFS) are collected by sensors and stored in the machinery fault database [6]. The database contains 1951 multivariate time series data comprised of six different simulated states such as

- Normal
- Horizontal misalignment
- Vertical misalignment

- Imbalance faults
- Underhung bearing fault
- Outer bearing faults

The rotatory machinery faults database contains the following percentage of each category of data as shown in Figure 6.1.

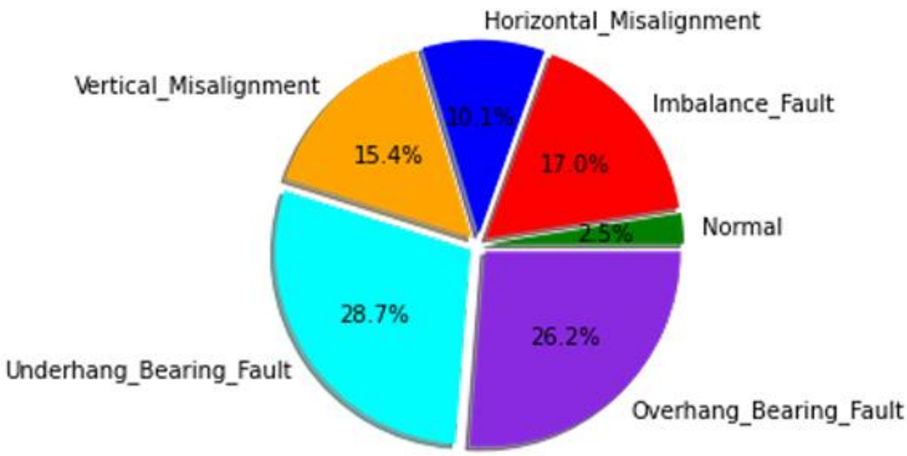

Figure 6.1: Rotatory machinery data percentage in machine fault database (MAFAULDA)

The rotatory machinery database contains the least amount of class normal data and maximum class under hang bearing faults data. The summary of the measurements is shown in Figure 6.2

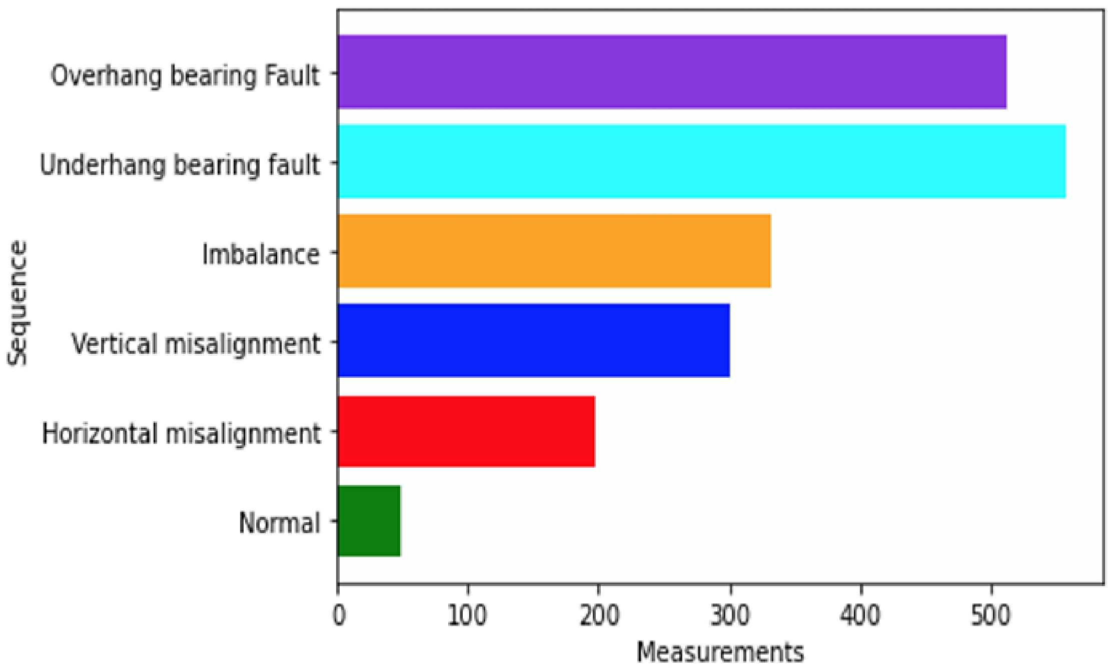

Figure 6.2: Summary of six states of rotatory machines measurement

6.4 Rotatory machine states

The data stored in the machinery fault database is acquired with the help of six accelerometers, a microphone, and a tachometer attached to the machine fault simulator [4]. It contains a total of 1951 scenarios as shown in Figure 4.2. The data describe the normal and five faulty states of the rotatory machine.

6.4.1 Normal

The normal sequence means without any fault. The 49 measurements of the normal sequence were used in this study. These sequences have been recorded with fixed rotation speed (range 737-3686 rpm) [6].

6.4.2 Imbalance

The total number of imbalance faults was 333 measurements [6]. The data was recorded with the load values (6g to 35g).

6.4.3 Horizontal misalignment

The number of horizontal parallel misalignment was 197 which was induced by each horizontal shift by the motor shaft shifting horizontally 0.5mm,1.0mm,1.5mm, and 2.0mm into MFS.

6.4.4 Vertical misalignment

The number of vertical parallel misalignment was 301 which was induced by each vertical shift by the motor shaft shifting horizontally 0.51mm,0.63mm,1.27mm,1.40mm,17.8mm, and 1.90mm into MFS.

6.4.5 Under hang bearing fault

In rotating machinery bearing is one of the most complex elements. Bearing faults are primarily causing failures in rotating machinery. When the bearing is placed between the rotor and motor in MFS. The under hang bearing fault has 558 total sequences with varying weights (0g, 6g, 20g, 25g).

6.4.6 Overhang bearing fault

When the rotor is placed between the bearing and motor in MFS. The overhang-bearing fault has 513 total sequences with varying weights (0g, 6g, 20g, 25g).

CHAPTER – 7

WORKING PRINCIPLE

Nowadays Artificial Intelligence (AI) has become popular in many other industries such as manufacturing and smart factories. The Internet of Things (IoT), Big Data (BD), and cloud computing make it more accessible to small industries as well. Machines in manufacturing industries have become smarter than before due to IoT, AI, and big data.

In recent times most of the manufacturing industries are transferred from preventive to predictive maintenance. This not only increases their productivity but also reduces cost. ML plays a significant role in such innovations. It also helps them to improve decision-making and accelerate discovery processes in manufacturing sectors. In the past different techniques have been used for industrial maintenance [4,5,7,8,22,23,24,25, 29,30].

In our study, we have used both classical machine learning and deep learning approaches to predict the fault in industrial machines as shown in the Figure 7.1. We followed the design of the Cross Industry Standard Process for Data Mining (CRISP-DM) model, which includes the following steps/processes;

- Business understanding: which includes the understanding of the industrial maintenance and their challenges and proposed solution.
- Data understanding: includes information/knowledge of our datasets.
- Data preparation includes the preprocessing steps that helped to prepare the data for downstream analysis.
- Modelling includes the steps where different analysis models and algorithm were applied.
- Evaluation includes the step where we evaluated the performance of the different models.
- Deployment includes our final model that was selected and applied to the data for the solution.

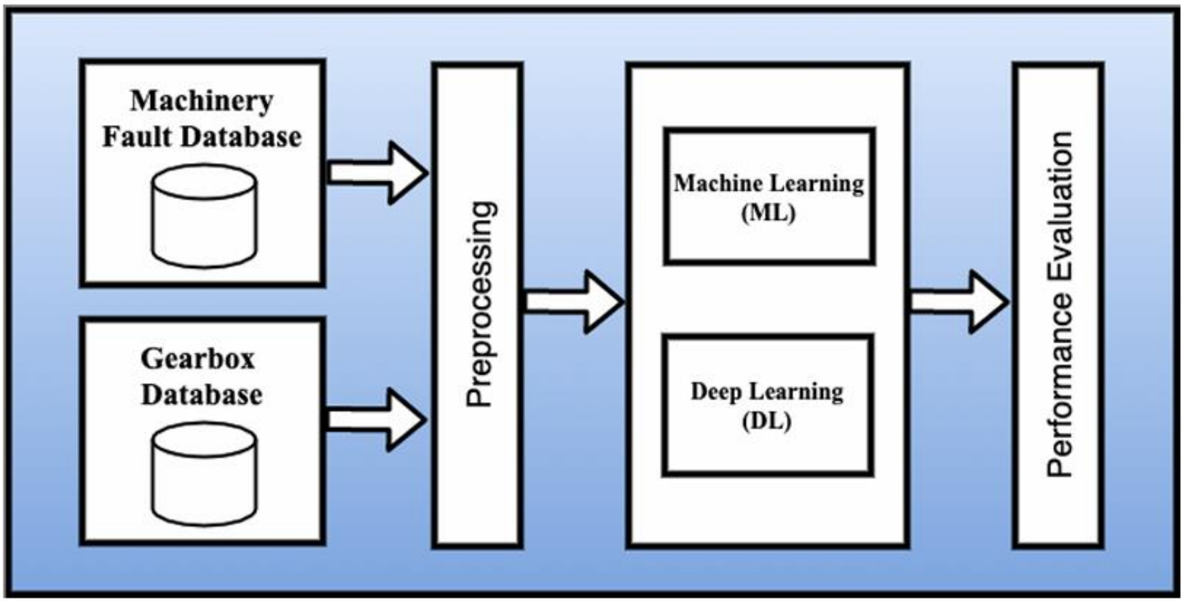

Figure 7.1: Machine learning and deep neural network pipeline for gearbox and rotatory machinery

7.1 Raw data / Sensors reading

The data from the gearbox have been collected by using four vibrations sensors as shown in Figure 7.2. The operating frequency used by sensors is 30Hz. These readings from the sensors are taken by varying load 0 to 90 percent and stored in the database. The gearbox database contains information about the health condition of the gearbox such as

- Broken teeth
- Normal

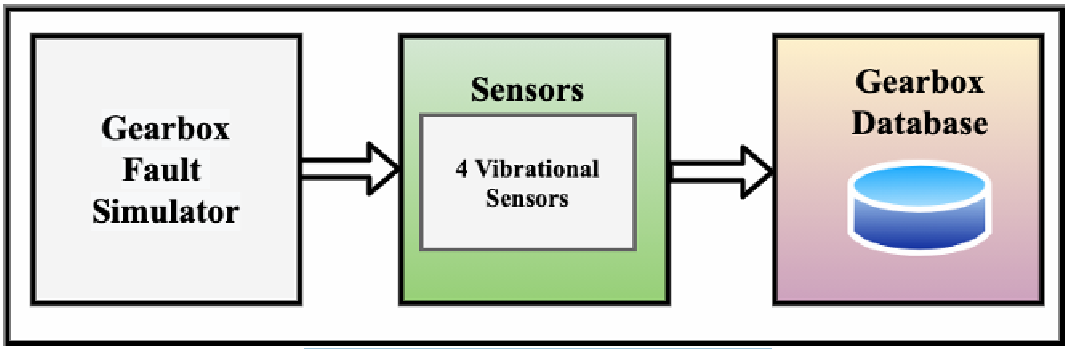

Figure 7.2: Gearbox data acquisition

The data stored in the machinery fault database is acquired with the help of six accelerometers, a microphone, and a tachometer attached to the machine fault simulator [4] as shown in Figure 7.3. It contains a total of 1951 scenarios with different operating conditions and loads. The data describe the normal and five faulty states of the rotatory machine.

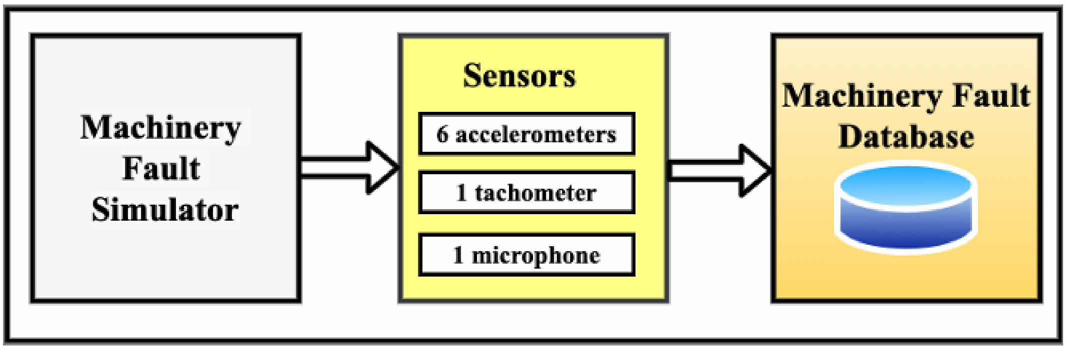

Figure 7.3: Rotatory machinery data acquisition

7.2 Preprocessing

It is an important step in any kind of analysis. During the preprocessing step, raw data is quality checked, trimmed, or cleaned to remove any bias in the data. The data coming from the databases is pre-processed by first doing the quality check where we check the missing (NAN) values. If the missing values are found, it is imputed with the mean value. In the next step the standard deviation of the dataset is performed and then labelled the data by categories (binary or multi class) specific to that dataset. Finally, the labelled dataset is merged into the single file containing all the required information. The data preprocessing helps us to

- Improve the quality of data
- Checking missing values
- Clean the data
- Normalized the data
- Transforming the data into the required format
- Find the outlier or noisy data before applying any ML or DNN model

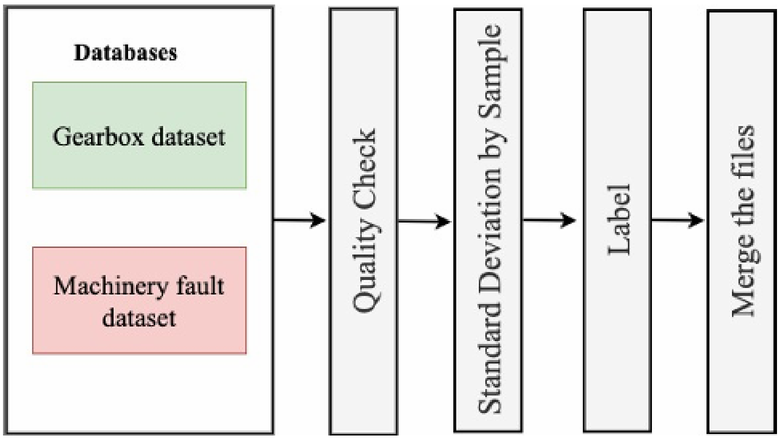

Figure 7.4: Preprocessing pipeline

7.3 Machine Learning Pipeline

In our study, we are dealing with the classification problem and our data are labelled so that is why we used supervised learning techniques. There are many supervised learning algorithms used to solve classification problems, but we used these algorithms

- Decision Tree
- Random Forest
- Adaboost (Adaptive Boosting)

When we applied the ML model to the gearbox and machinery faults study, our initial goal was to learn and test the different types of ML algorithms. Therefore, we selected only those algorithms that minimized the type 1 and type 2 errors as minimum as possible. Another reason for using decision trees and random forests was that they can be used for classification and regression problems.

7.4 Deep Neural Network (DNN) Pipeline

The goal of using DNN pipeline is to improve the efficiency of the model on given datasets. Relu activation function is used at the input, and hidden layers and sigmoid is used at the output layer as shown in the Figure 6.5. Different neurons were used in each layer. This combination of neurons given us the desired results.

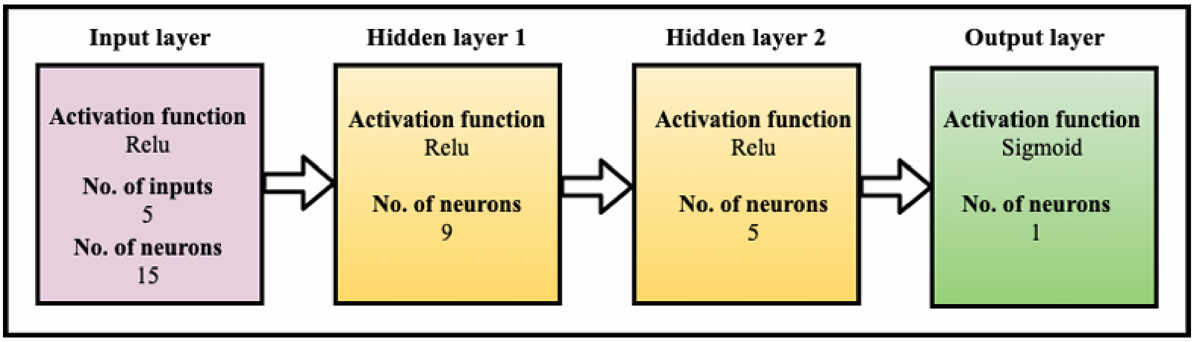

Figure 7.5: DNN architecture for gearbox

The DNN model of rotatory machines contains two hidden layers, one input and output layer. Relu activation function is used in the input, and hidden layers and SoftMax is used at the output layer as shown in Figure 6.5.

CHAPTER – 8

RESULTS

Training and test data are required to build and validate the results of the machine learning (ML) and deep neural network (DNN) model. Here we have analysed two datasets i.e., gearbox and machinery fault studies, which are further divided into training and test datasets to build and evaluate the performance of ML and DNN models. The models are learned from the training set and performance is evaluated on test data or unseen data. In both studies, seventy percent of the data is used for training and thirty percent is used for testing the models as shown in Figure 8.1.

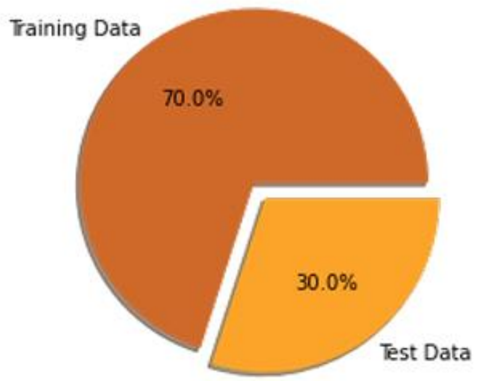

Figure 8.1: Ratio of training and test data

8.1 Gearbox Fault Prediction

The gearbox fault prediction dataset (n=4000000) consists of only two classes: normal and broken teeth. It is a binary classification problem. The training data contains 2800000 records (70 %) and the test data contains 120000 records (30 percent). The records are equally distributed among the classes. This means we have a balanced classification problem Figure 8.2.

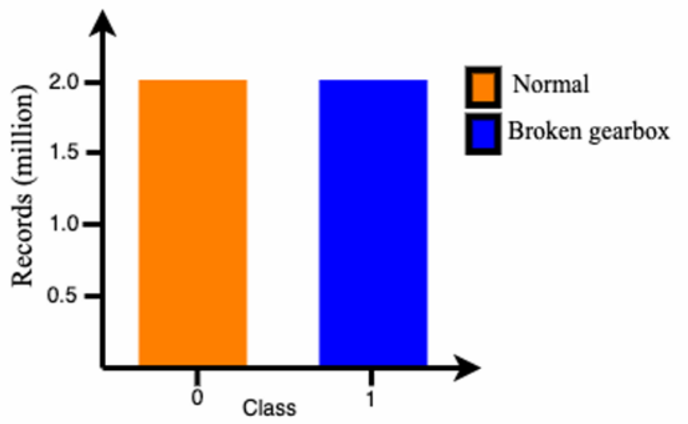

Figure 8.2: Data distribution among normal and broken gearbox classes

The information from the descriptive features was obtained with the help of sensors. All the readings from the sensors were numerical. This is a binary classification problem, where 0 means normal class and 1 means broken gearbox teeth class. We have 5 descriptive and one target feature.

8.1.1 Performance Evaluation on raw data

Figure 7.3 describes the results of ML and DLL models on the raw data of the gearbox dataset. We first evaluated and compared the performance of different machine learning models ML (Figure 8.3 (a-c)) and DL model (Figure 8.3d) using the gearbox raw data i.e. without applying any normalization techniques. It means that models were first directly deployed on raw data. Our results showed that type 1 error was highest in the random forest (RF) (Figure 8.3c) and type II was lowest. The DNN model showed that the type 1 error was lowest while the type II error was highest (Figure 8.3d).

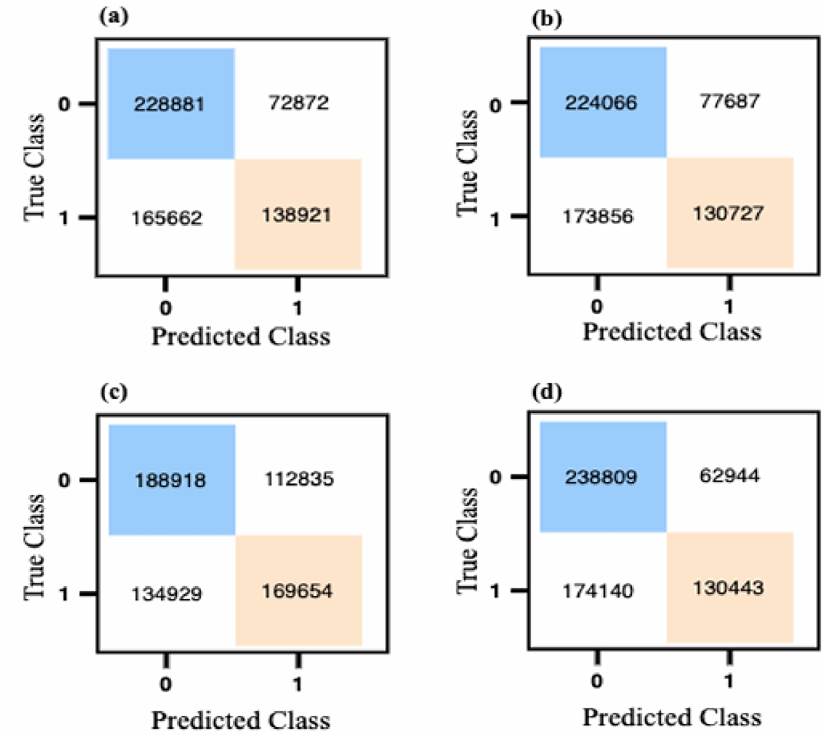

Figure 8.3. Performance evaluation of ML and DLL models on raw data

In each confusion matrix predicted class (x-axis) and true class (y-axis) representing the normal class with ‘0’ and broken gearbox tooth class ‘1’. Confusion matrix of (a) decision tree (b) AdaBoost (c) random forest (d) deep neural network.

Overall performance of DNN and decision tree is better based on highest accuracy rate/precision rate and lowest error rate than the random forest and AdaBoost. However, the F1 score was best for the RF model.

We next compared the graphical description of the ML and DNN classifier performance on raw data using the ROC. Our results show that the area under the ROC curve of the DNN model is higher than the ML model (RF, DT, and AdaBoost) models (Figure 8.4). Based on the ROC curve and AUC score performance of DNN and Random Forest is much better than the decision tree and AdaBoost (Figure 8.4 a-d).

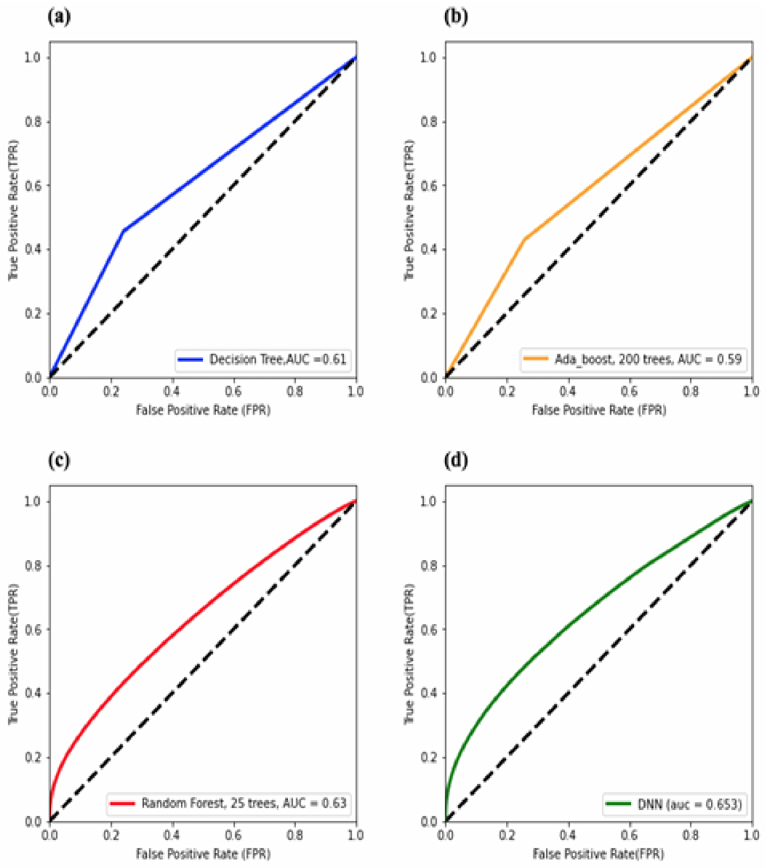

Figure 8.4. ROC curve and corresponding AUC score of both ML and DNN model on raw data

8.1.2 Performance Evaluation of normalized data

The standard deviation with sample size (N=10,25,50,100 and 500) used to normalized the gearbox and machinery fault datasets. The motivation for using these sample values was to reduce the error rate and increase the performance of ML and DNN model.

- Normalized data with sample size N=10
  We next evaluated the performance of different models using the normalized data by taking the sampling size of N=10. Here instead of directly taking the raw data from the sensor reading, we first take standard deviations for each of the sample sizes 'N=10', and then the models were deployed on this normalized dataset.

Based on this approach, our results showed that the overall performance of all the models, both the ML and DNN was improved by approximately 10% as compared to the raw data (Figure 8.5 (a-d)). The accuracy rate of the DNN model was improved from 60% in raw data to 73% in normalized data.

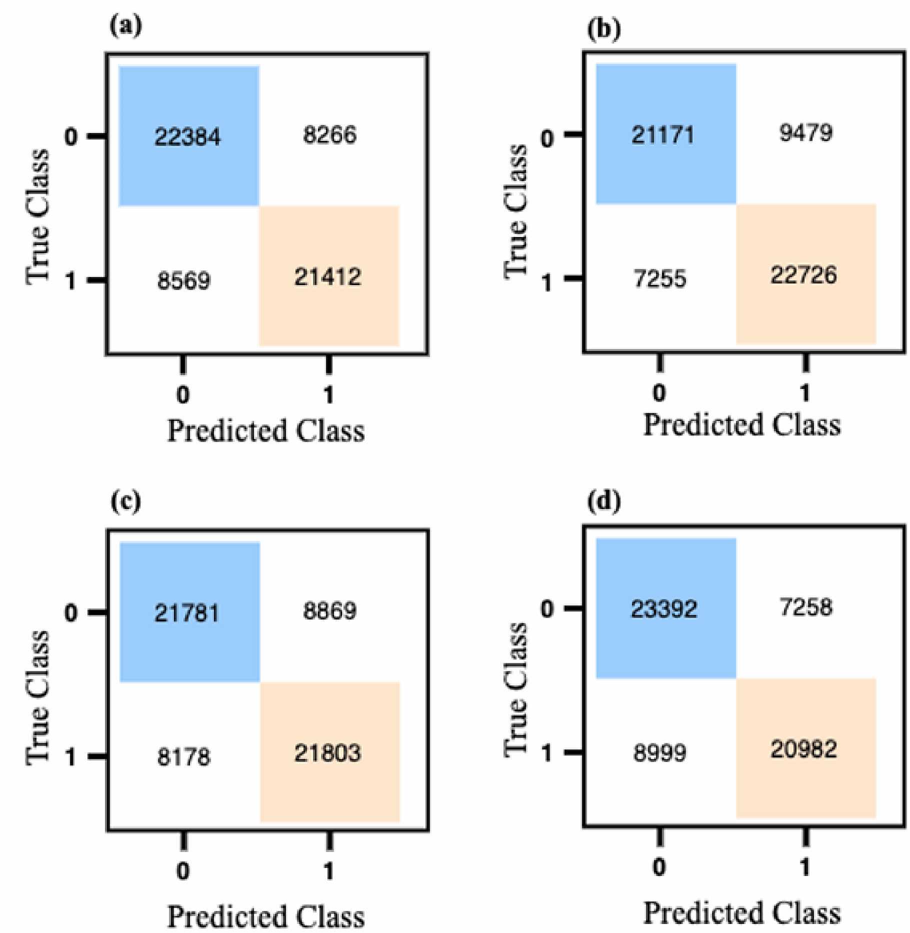

Figure 8.5. Performance evaluation of ML and DLL models with N=10

The ROC and AUC were significantly improved for all the models using the normalized data with the sampling size N=10 as compared to raw data (Figure 8.6(a-d). Interestingly, we observed that the AUC for the DNN model improved to 0.82 in normalized data compared to raw which was 0.65.

- Normalized data with sample size N=100
  The performance of the ML and DNN models was further evaluated by taking the normalized data with the sampling size of N=100. The overall performance of all models improved remarkably, with the DNN and RF models showing the best results as can be seen in the confusion matrixes (Figure 8.6). The accuracy rate of the DNN and RF model 33 reached approx. 93% with the highest precision and F1-scores and the lowest error rate of about 7%.

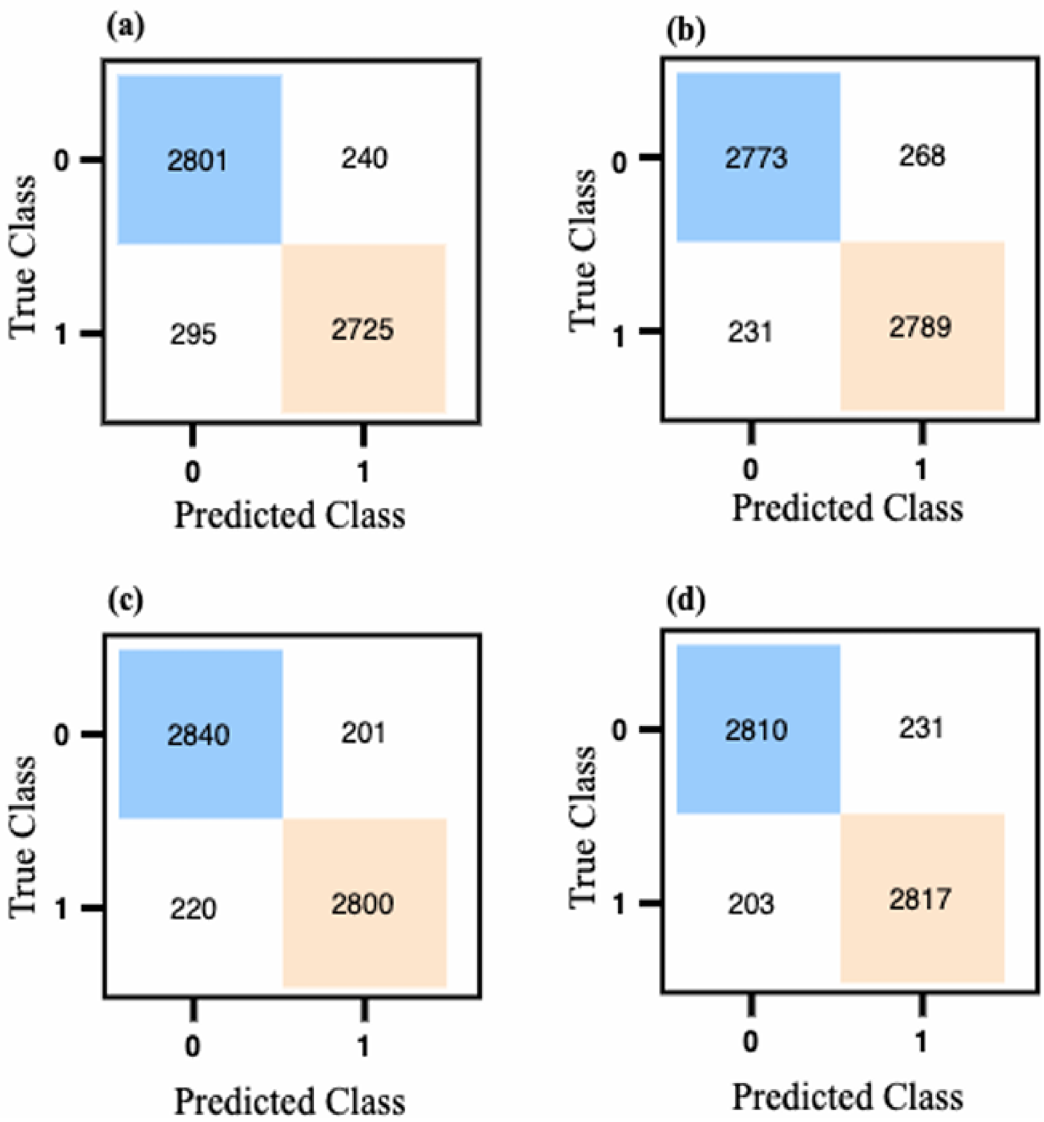

Figure 8.6. Performance evaluation of ML and DLL models with N=100

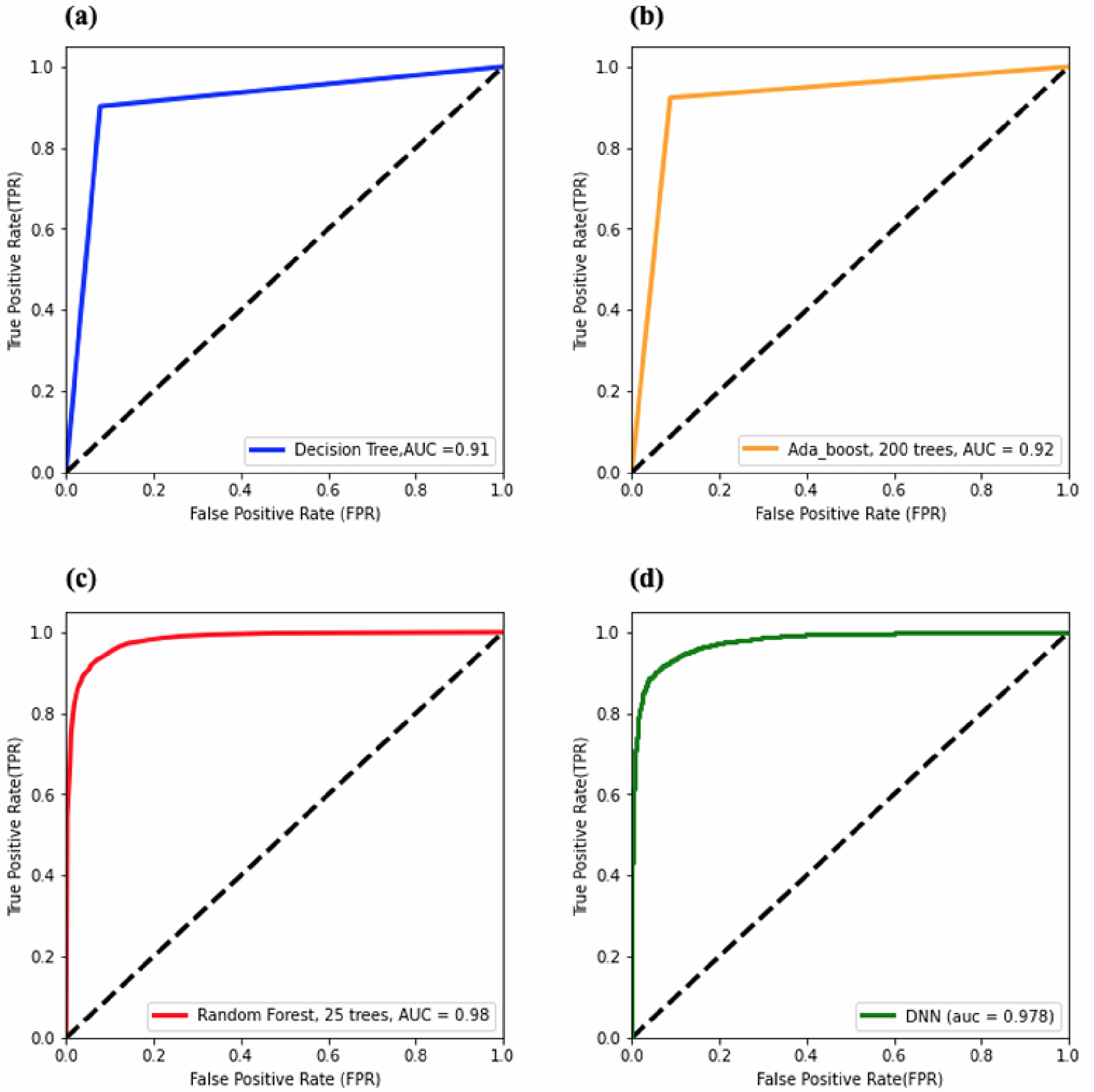

Figure 8.7. ROC curve and corresponding AUC score of both ML and DNN model with N=100

The ROC further improved for all models in this case and the AUC of both RF and DNN models was about 0.98. As previously, the RF and DNN models' performance was very similar using sample size N=100.

8.2 Machinery Fault Prediction

The second dataset that we analysed here was the machinery fault prediction (MFP) dataset, which is a multi-classification problem. We have six classes such as normal, imbalance, horizontal misalignment, vertical misalignment, under hang, and overhang bearing faults in this study. The training data contains 6828550 records (70 %) and test data holds 292650 records (30 percent). The distribution of records among the multi-classes in the MFP dataset is shown in Figure 8.8.

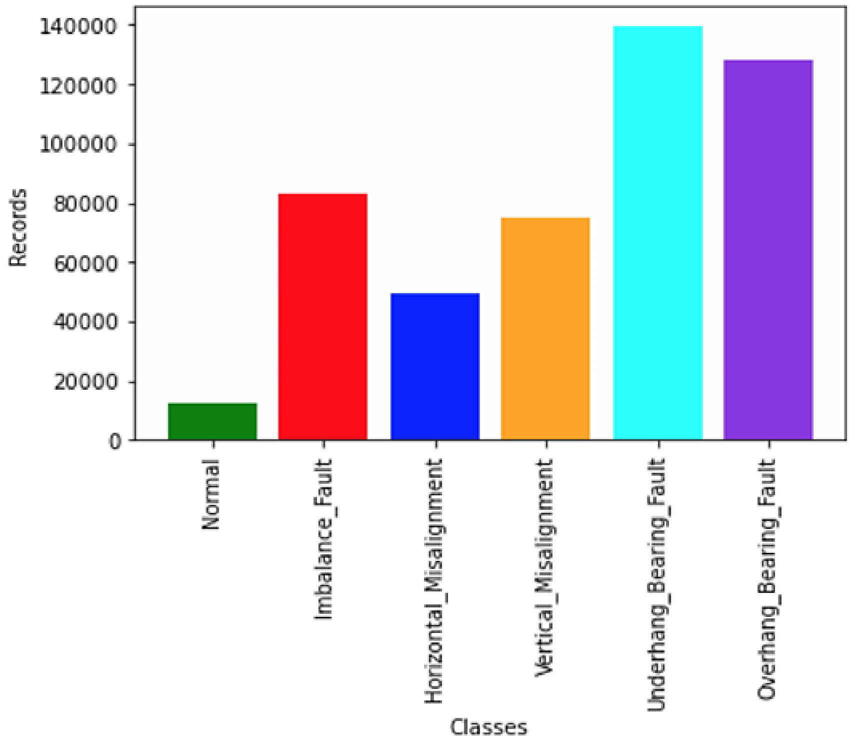

Figure 8.8. The number of records in each of the six classes in MFP

The records from the descriptive features were obtained with the help of sensors. All the readings from the sensors are numerical. This is a multi-classification problem, where '0' represents normal class, '1' represents imbalance fault class, '2' represents horizontal misalignment, '3' represents vertical misalignment, '4' represents under hang bearing fault and, '5' represents overhang bearing fault classes. We have 8 descriptive and one target feature.

The results of both ML and DNN models were evaluated and compared with each other. The performance of the models is evaluated by confusion matrix, accuracy, F1-score, AUC score, and ROC curve. The error was calculated by using MSE.

8.2.1 Performance evaluation of ML model in MFP dataset

This section describes the performance of the ML model on the given datasets. The algorithm we used in ML is random forest. The confusion matrix is shown in (Figure 8.9). summarizing the performance of the model. The correctly classified classes using this model were shown diagonally in the confusion matrix (Figure 8.9). While other elements (non-diagonal) of the confusion matrix indicate incorrectly classified records. Our results showed that 153 cases of class normal were incorrectly classified into other classes such as 11 records were classified as class imbalance, 90 records in horizontal misalignment, 22 records in vertical misalignment, 13 records in under hang bearing fault, and 17 records in overhang bearing fault (Figure 8.9). The confusion matrix helps to analyse different types of errors in classification.

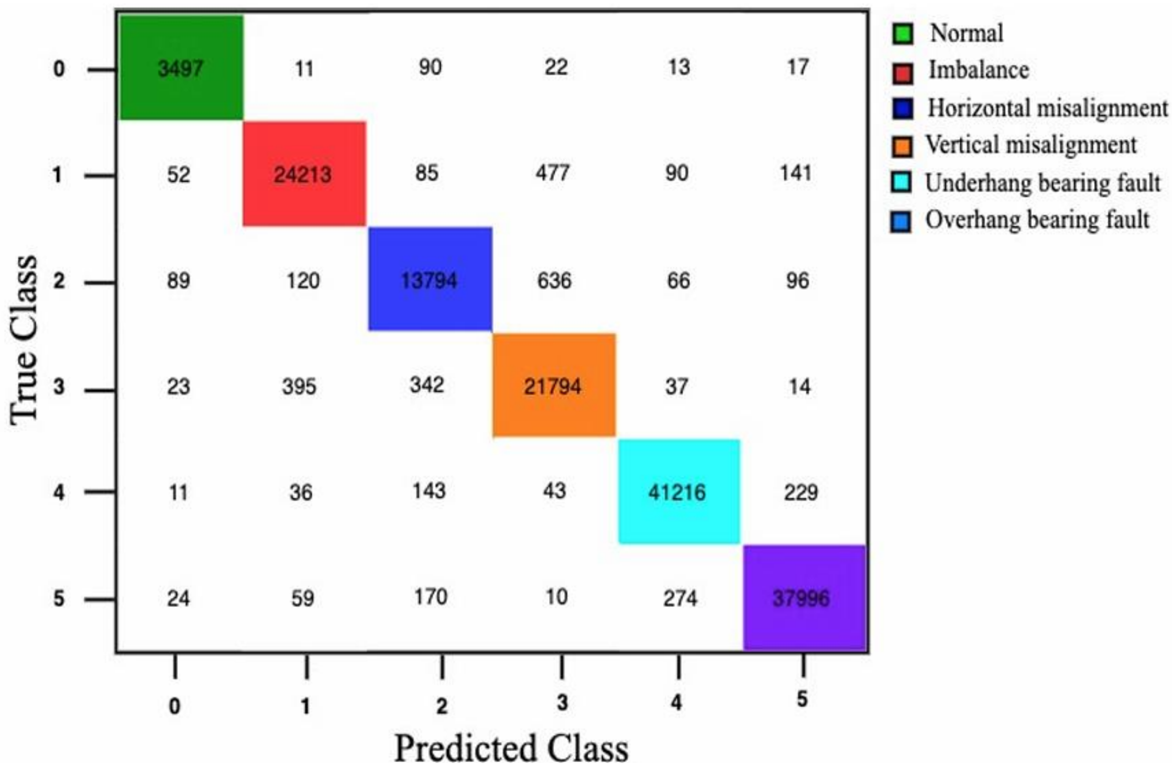

Figure 8.9. Performance evaluation of RF of MFP using a confusion matrix

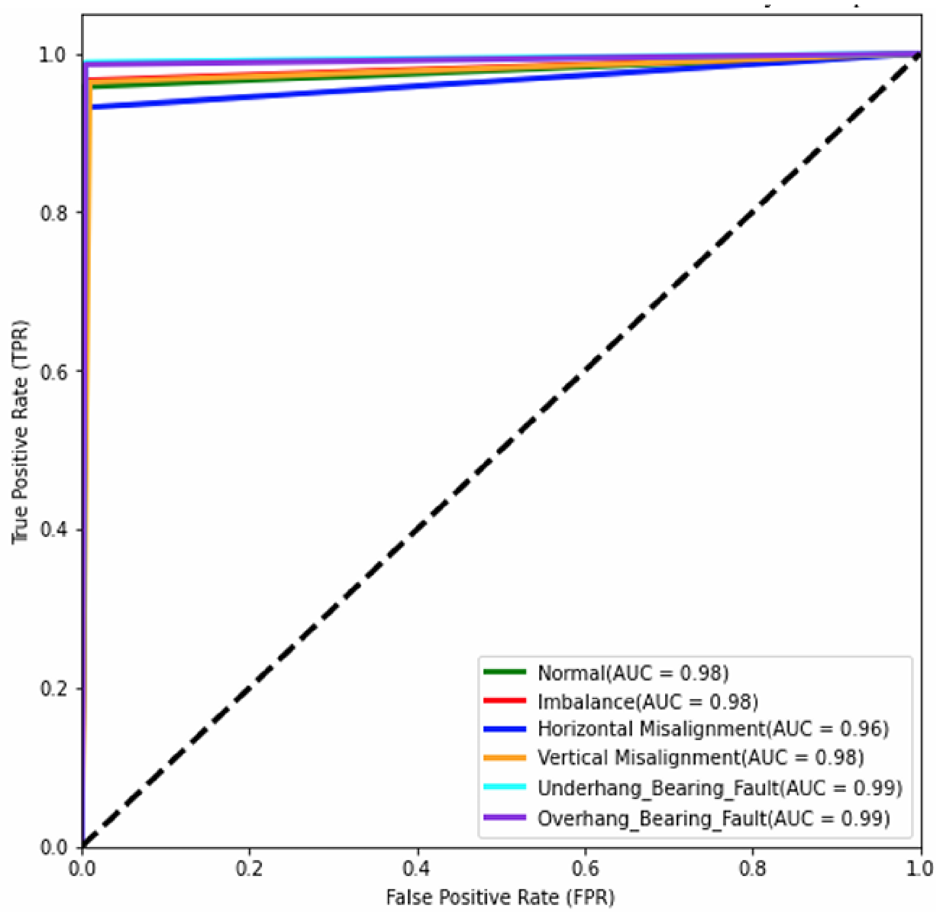

Figure 8.10: ROC curve and corresponding AUC score of RF model in MFP dataset

The under hang bearing fault AUC score of 0.99 remained the same as in the RF model. While all other classes, AUC has slightly decreased compared to the RF model, showing that the RF model performed better than the DNN model for the MFP dataset (Figure 8.10).

In the DNN model, we have also compared the predicted output to the actual output. Here the error was calculated between actual and predicted output. During backpropagation of DNN model weights are updated with each iteration or epochs. The error is reduced, and accuracy is increased with each iteration as shown in Figure 8.11.

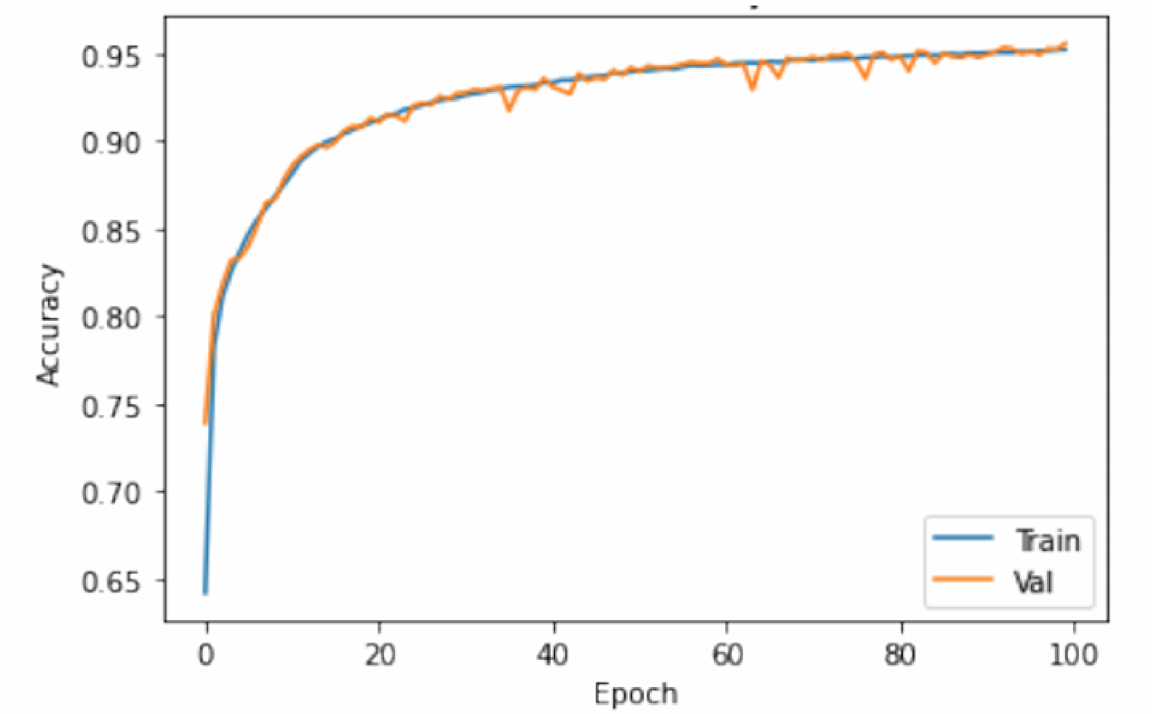

Figure 8.11: Epoch vs Accuracy using the DNN model in MFP dataset

The test loss was reduced, and the accuracy was improved when the number of epochs is increasing (Figure 8.11).

CHAPTER – 9

DISCUSSION

9.1 Gearbox Fault Prediction

To predict the gearbox fault with ML/DNN is very challenging. As it requires a huge amount of historical data with varying equipment operating conditions to build these models. Once the data is taken from the equipment with the help of sensors and stored in the database. This data is used for training and testing the ML/DNN model. Sometimes these data required different types of preprocessing techniques before deploying the ML model. We compared the performance of our models with and without applying any preprocessing or normalization techniques on this dataset.

In the case of the gearbox, we have a binary class problem with two classes: normal and broken teeth gearbox. The data among the classes were equally distributed. So, that the classes are balanced. It is a classification problem and data are labelled. We applied supervised machine learning techniques. This dataset has five descriptive and one target feature. Seventy percent of data is used for training and thirty percent is used for testing as shown in the figure.

The performance of our ML/DLL models was not good on the raw gearbox dataset when we deployed the models on raw data without any preprocessing or normalization techniques. The accuracy, F1 score was very low. AUC score was below 0.65 and the models had difficulty differentiating between normal and broken gearbox classes. Although we have tested different ML algorithms on the raw data, the result was not much improved. The MSE error was also high. The reason for very low accuracy, an F1 score, and a very high error rate of around 40 % could be due to noise and external environment from the sensor readings. The purpose of using these techniques was to get the desired results without applying any preprocessing or normalization techniques to reduce the computational cost.

However, the performance of the ML and DNN models gradually improved when we took the sample standard deviation of sensor readings. In this case, we have used different sampling sizes such as N=10,25,50 and 100. When the sampling size has increased the accuracy, F1-score, AUC score was also increased, and the error rate was decreased. Overall, the performance of the models was significantly improved as compared to applying these models without any sampling.

This helped in removing any noise and we got desired results by gradually increasing the sample size. The accuracy and F1-score were also improved at each preprocessing and normalization step. The overall AUC score was improved to 0.98 with an average accuracy of 93%. The model easily differentiated between normal and broken gearbox classes with an AUC score of 0.98.

The algorithms were ranked based on their performance such as accuracy, error rate, F1-score, AUC score, and ROC curve. Overall performance of DNN model and random forest were very good as compared to decision tree and AdaBoost on this dataset. The ROC curves of these models became very smooth. AUC and F1 scores were also very high, and the error rates were very low. The DNN model was ranked first based on our results and the decision tree ranked last. Hence, I will suggest deploying DNN and random forest models on this type of dataset to get the desired results.

One of the drawbacks of this gearbox fault dataset was that we have classified only two types of gearbox conditions such as normal and broken teeth, but we didn't have any data and information 41 about the other gearbox faults such as gearbox misalignment, backlash, etc. Another limitation of this dataset was that the data taken from the simulator were recorded with predefined conditions instead of the fault occurring randomly.

9.2 Machinery Fault Prediction

Industrial machines are composed of both electrical and mechanical components. The prediction of fault in the mechanical components is challenging. Because in a single machine there are a lot of mechanical components such as gearbox, bearing, shaft, rotor, etc. You will need different types of sensors to measure and observe the behaviour of each mechanical component.

In the case of the machinery fault prediction dataset, the data used to build the ML and DNN model to predict the machinery fault were from the spectra quest machinery fault simulator. Unlike gearbox fault prediction, here we have multi classes such as Normal, Imbalance, Horizontal misalignment, Vertical misalignment, under hang bearing fault, Overhang bearing fault.

The ML model is built with random forests and the performance was evaluated using the confusion matrix, accuracy, F1- score, and ROC in ML model on this dataset. The ratio of correct prediction was more than 90 percent. Accuracy, F1-score, AUC-score were also very high. MSE was very low.

Once the data is acquired from the database. The quality of data was checked, and then sample standard deviation with sampling size (N=500) was applied to the sensor's reading. This helped us to minimize any error from the sensor reading and remove any noise.

The performance of the ML and DNN models on machinery fault prediction datasets was almost the same for all the classes except the normal class. The area under the ROC curve of class normal is reduced from 0.98 to 0.95 in the DNN model. Both normal and horizontal misalignment classes are imbalanced data among other classes. So that is why their AUC scores are slightly low as compared to other classes.

One of the drawbacks of the machinery fault prediction dataset was that the data taken from the simulator were recorded with predefined conditions. The normal and horizontal misalignment classes data was very small as compared to the different faults. Although we classified different types of machinery faults in this study, in the case of bearing faults we classified only two types such as under hang and overhang faults. It would have been nice to investigate the subtype of bearing faults such as ball, rolling, and outer faults as well. This would have helped the maintenance team to know the exact type of bearing fault instead of general faults. Another limitation of this dataset was that there was a lack of data about the broken bearing and other mechanical components to build the model.

CHAPTER – 10

CONCLUSION

In this study performance of machine learning (ML) and deep neural network (DNN) were compared and evaluated on gearbox and machinery fault datasets. In ML we used different algorithms such as decision tree, random forest (RF), and AdaBoost to build the model. Overall, the performance of the random forest is very good as compared to the decision tree and AdaBoost. The DNN model also performs well on both datasets, but the biggest challenges faced to build these models were the selection of hyperparameters, several hidden layers, activations functions, and loss functions to get the desired results. Our results show RF and DNN models have better fault prediction ability to identify the different types of machinery and gearbox faults as compared to the decision tree and AdaBoost. In the future, we need to investigate statistical and recurrent neural network approaches as well. Especially we need to study autoregressive integrated moving average (ARIMA) and long short-term memory (LSTM) models. The hybrid approach, which is a combination of statistical models with ML, DL, LSTM, RNN models will be very helpful in predicting missing data from the sensors. One of the challenges of predicting faults in industrial machinery is that you require a lot of historical data to build the ML models. Industrial machines are operated in different conditions and getting the data from each component of the machine is also tricky, you require a resource to record the data from the equipment and store it in a cloud or particular place. The biggest challenges of implementing these approaches in industries are currently IoT-based devices are only affordable for big companies and manufacturing units to monitor their equipment. We need to investigate how these ML-based predictive approaches can be transferred to small companies as well. So that they can benefit from artificial intelligence.

FUTURE SCOPE

The future scope for predictive maintenance in machine tools is quite promising, driven by advancements in technology and industry demands for efficiency and cost-effectiveness. Here are some aspects contributing to its future growth:

Advanced Analytics and AI

As AI and machine learning algorithms continue to improve, they will become more adept at predicting equipment failures with greater accuracy. This will enable maintenance teams to address issues before they become critical, reducing downtime and costs.

IoT Integration

The Internet of Things (IoT) plays a crucial role in predictive maintenance by enabling real-time monitoring of machine performance and health. Integrating machine tools with IoT sensors allows for continuous data collection, analysis, and proactive maintenance scheduling.

Big Data Handling

With the proliferation of sensors and connected devices, the volume of data generated by machine tools is increasing exponentially. Future advancements in big data technologies will enable more efficient processing, storage, and analysis of this data, leading to more accurate predictive maintenance models.

Edge Computing

Edge computing brings processing power closer to the data source, reducing latency and enabling faster decision-making. In the context of predictive maintenance, edge computing allows for real-time analysis of machine data, enabling immediate action to be taken to prevent failures.

Augmented Reality (AR) and Virtual Reality (VR)

AR and VR technologies can enhance the maintenance process by providing technicians with immersive training, remote assistance, and visualizations of equipment data. These technologies can improve troubleshooting efficiency and reduce the need for on-site visits.

Digital Twins

Digital twins are virtual replicas of physical assets that can simulate their behaviour in real-time. By creating digital twins of machine tools, operators can predict and simulate potential failures, optimize performance, and test maintenance strategies in a risk-free environment.

Predictive Maintenance as a Service (PMaaS)

As predictive maintenance technologies mature, we may see the emergence of PMaaS offerings, where third-party providers offer predictive maintenance solutions as a subscription service. This could make predictive maintenance more accessible to smaller businesses that may not have the resources to develop and maintain their own systems.

Regulatory Compliance and Industry Standards

Regulatory requirements and industry standards regarding equipment maintenance and safety are continually evolving. Future predictive maintenance systems will need to adapt to these changes to ensure compliance and maintain operational integrity.

REFERENCES

[1] “Why Implement a Predictive Maintenance Strategy?,” Vimana. https://govimana.com/predictive maintenance-strategy/ (accessed Apr. 4, 2021).

[2] “Machine Learning Techniques for Smart Manufacturing: Applications and Challenges in Industry 4.0,” Intellectyx, Jul. 09, 2019. https://www.intellectyx.com/blog/machine-learning-for-smart manufacturing/ (accessed Jul. 12, 2021).

[3] “What are the types of industrial maintenance and how can IIoT help?,” Netilion Blog, Sep. 02, 2019. https://netilion.endress.com/blog/types-industrial-maintenance-iiot/ (accessed Apr. 9, 2021).

[4] F. Ribeiro, M. Marins, S. Netto, and E. Silva, “Rotating machinery fault diagnosis using similarity-based models,” presented at the XXXV Simpósio Brasileiro de Telecomunicações e Processamento de Sinais, 2017. doi: 10.14209/sbrt.2017.133.

[5] A. Alzghoul, A. Jarndal, I. Alsyouf, A. A. Bingamil, M. A. Ali, and S. AlBaiti, “On the Usefulness of Pre-processing Methods in Rotating Machines Faults Classification using Artificial Neural Network,” Journal of Applied and Computational Mechanics, Jan. 2021, doi: 10.22055/jacm.2020.35354.2639.

[6] “MAFAULDA: Machinery Fault Database [Online].” http://www02.smt.ufrj.br/~offshore/mfs/page_01.html (accessed May. 22, 2021).

[7] “Classification of faults in gearboxes — pre-processing algorithms and neural networks | SpringerLink.” https://link.springer.com/article/10.1007/BF01413861 (accessed Aug. 23, 2021).

[8] "Gearbox fault diagnosis using data fusion based on self-organizing map neural network - Zhang Qiang, Gu Jieying, Liu Junming, Tian Ying, Zhang Shilei, 2020." https://journals.sagepub.com/doi/full/10.1177/1550147720923476 (accessed Aug. 23, 2021).

[9] "5 Types of Industrial Maintenance Programs," Quality Millwright, Jun. 19, 2019. https://www.qmillwright.com/5-types-of-industrial-maintenance-programs/ (accessed Apr. 17, 2021).

[10] M. Bentley, "Machine Learning for Predictive Maintenance - Top Opportunities for 2020-2021 | The Ritz Herald," https://ritzherald.com/. https://ritzherald.com/machine-learning-for-predictive-maintenance/ (accessed May. 17, 2021).

[11] "Disadvantages of doing Preventive Maintenance," Bayt.com. https://specialties.bayt.com/en/specialties/q/95601/disadvantages-of-doing-preventive-maintenance/ (accessed May. 21, 2021).

[12] "Predictive Maintenance – What you need to know | IoT Now News & Reports," IoT Now News - How to run an IoT enabled business, May 02, 2018. https://www.iot-now.com/2018/05/02/81526 predictive-maintenance-need-know/ (accessed Jul. 23, 2021).

[13] "What Are Benefits and Drawbacks of Preventive Maintenance?" https://www.onupkeep.com/answers/preventive-maintenance/benefits-of-preventive-maintenance (accessed Jun. 29, 2021).

[14] "What is Predictive Maintenance? [Benefits & Examples]," Fiix. https://www.fiixsoftware.com/maintenance-strategies/predictive-maintenance/ (accessed Apr. 7, 2021).