DESIGN AND DEVELOPMENT OF A GENERIC PROGNOSTICS SIMULATOR

M.Tech. Thesis

By

SHUBHAM B. MENKAR

DISCIPLINE OF MECHANICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

JUNE 2020

DESIGN AND DEVELOPMENT OF A GENERIC PROGNOSTICS SIMULATOR

A THESIS

Submitted in partial fulfillment of the

requirements for the award of the degree

of

Master of Technology

by

SHUBHAM B. MENKAR

DISCIPLINE OF MECHANICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY

INDORE

JUNE 2020

INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **DESIGN AND DEVELOPMENT OF A GENERIC PROGNOSTICS SIMULATOR** in the partial fulfilment of the requirements for the award of the degree of **MASTER OF TECHNOLOGY** and submitted in the **DISCIPLINE OF MECHANICAL ENGINEERING Indian Institute of Technology Indore** is an authentic record of my own work carried out during the time period from July 2018 to June 2020 under the supervision of **Dr Bhupesh Kumar Lad**, Associate Professor, (PhD).

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

29/06/2020

Signature of the student with date

(Shubham Balmukund Menkar)

This is to certify that the above statement made by the candidate is correct to the best of my/our knowledge.

Signature of the Supervisor

(Dr. B.K. Lad)

Shubham Menkar has successfully given his M. Tech. Oral Examination held on **25/06/2020**.

Signature of Supervisor of M. Tech. thesis
Date: 30/6/20

Convener, DPGC
Date: 01/07/2020

Signature of PSPC Member
Date: 01/07/2020 (Dr. Pavan K. Kankar)

Signature of PSPC Member
Date:01/07/2020 (Dr. Anand Parey)

ACKNOWLEDGEMENTS

I take this opportunity to express my profound sense of gratitude and respect to all those who helped me through in the completion of this work. I would first like to thank my thesis supervisor Dr Bhupesh Kumar Lad, for his guidance and support. Professor helped me a lot whenever I ran into a trouble spot or had a question about my research or writing. He provided me with the freedom to work on this project but steered me in the right direction whenever he thought I needed it. His trust in me pushed and inspired me during the completion of my work. I specially thanks to Mr. Ram Mohril for his consistent encouragement and guidance. Also, special thanks to Mr. Kushagra Sharma for helping with programming related work in my project. Finally, I am thankful to my friends and colleagues here at IIT Indore for their help and encouragement.

Shubham Menkar
M. Tech. (**Production & Industrial Engineering**)
The discipline of Mechanical Engineering
IIT Indore.

DEDICATION

I dedicate this thesis

to

My beloved family

My friends

and

My guide.

ABSTRACT

Industries are transforming in the fourth industrial revolution, from traditional manufacturing and industrial practices to smart manufacturing with the assistance of modern smart technology. Prognostics is a key science in industry 4.0 which assists in condition-based maintenance of industrial assets. Hence researchers, scientists and different industrial organisations are finding different ways to implement prognostics effectively. Prognostics has evolved as an essential tool for the estimation of the Remaining Useful Life of the critical components of several industrial assets.

Prognostics requires a large amount of life data of the component or system for the accurate estimation of the remaining useful life of that component or system. However, the data collection process used for prognostics is a cumbersome process which consists of operating the asset from healthy condition to failure. It is not advisable to run the asset until failure because industrial assets are expensive. Also running the asset till failure is a destructive and time-consuming process. This situation creates a problem for the data collection process in prognostics and health management.

This thesis focuses on resolving the scarcity of industry-grade prognostics data. For this purpose, a mechanism which can generate prognostics data without need of running the actual machine has been designed and developed. This mechanism uses the historical data of the machine for which the user wants to generate data. This historical data is condition monitoring data of that machine.

To resolve the problem of data scarcity, a mechanism is required, which can provide the prognostics data similar to the actual data. For this purpose, this thesis proposes a solution named as a Generic Prognostics Simulator. GPS uses the simulation process for new data generation. Primary factors considered while developing this mechanism is; its

ability to generate the degradation data of several mechanical components. This simulator uses the historical data of asset to generate new data sets using several features of that data. A generic algorithm is developed for the processing of GPS, i.e. to simulate the datasets.

The basic need for massive prognostics data can be satisfied using GPS without running the actual component until failure. Also, industrial organisations can save a large amount of cost and time using GPS for the data generation process

GPS enables data generation from historical condition monitoring data which resolves massive data requirement of industries. It also assists in training the latest prognostics models developed by academic researchers on the industry-grade data generated by GPS. Many research students struggle generating new models for prognostics due to lack of data; GPS can resolve this problem.

The validation of the data generation process with the help of GPS using historical condition monitoring data consisting of wear and vibration data of a milling cutter has been done. The detailed results are discussed in chapter number five. Results show the similarity between actual datasets and simulated datasets which is illustrated using graphical representations of simulated data with the original datasets.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AWV: Analogue Write Value

CBM: Condition Based Maintenance

CIM: Change In Mean

CMD: Condition Monitoring Data

DC: Direct Current

ERP: Enterprise Resource Planning

PHM: Prognostics and Health Management

RMS: Root Mean Square

RMSE: Root Mean Square Error

RPM: Rotation Per Minute

RSM: Randomly Started Mean

RTF: Run To Failure

RUL: Remaining Useful Life

SD: Sample Datasets

TS: Time Stamp

Chapter 1. INTRODUCTION

Industries are turning towards the fourth industrial revolution. It requires technologies such as prognostics implemented in the industry. However, a large portion of the industries is facing several problems while implementing prognostics.

Prognostics uses system information for estimation of Remaining Useful Life (RUL), i.e. either it uses physical properties of the system for predictions, or it uses Condition Monitoring Data (CMD) of the system. It has been noticed that the data collection process is costly and time-consuming. This is the reason for the scarcity of industry-grade prognostics data and unavailability of the data for prognostics is one of the significant problems in the implementation of prognostics. Due to this problem, many industries are not implementing prognostics. Hence the data collection process becomes very crucial for prognostics and to resolve the problem of unavailability of a large amount of prognostics data is a significant challenge.

Thesis work identifies the problem of data scarcity and proposes a solution to the problem. The solution fulfils the requirement of data in the industry. Also, the generation of new datasets can ease the generation of new prognostics models to researchers.

1.1 Prognostics and Health Management (PHM)

PHM is a tool which provides solutions for maintaining the system health at the maximum possible level and prevent any breakdowns. A prognostics algorithm is implemented where a component degrades with respect to operating time and fails after reaching the user-defined threshold of degradation parameter. This degradation data is used to predict the RUL of that component. The historical data is used by PHM for extrapolation of the current trend of degradation and predicting possible failure time in the future. Maintenance is a process of preserving the asset health condition. The predicted RUL is an essential concept in decision making for maintenance activities and to prevent accidental conditions.

Prognostics provides information about the current health and expected future degradation of the asset used for effective asset management. This information can be collected from the diagnostics and prognostics system, respectively [1]. PHM is gaining importance in industry and academia.

Diagnostics and prognostics technologies were enabled in the mechanical industry, as maintenance technology came into light. Implementing maintenance proactively directly reduced the cost of maintenance as well as increased machine availability. Prognostics assists in scheduling the maintenance exactly before the failure, with the use of Condition Based Maintenance (CBM). To achieve this, conventional fail and fix maintenance strategy must be switched to predict and prevent. CBM is one of the modern technologies that falls well within the framework of PHM. (Lee, et al., 2014) highlights that the use of prognostics for estimation of RUL is assisting the maintenance activities for the last fifteen years [2].

PHM provides the user with information about the fault generated in the component in an early stage. PHM also assist in monitoring and forecasting the severity of the progression of the fault, and to help in the planning and autonomously triggering maintenance schedule. Proper planning of maintenance activities helps in managing the inventory levels of required components for maintenance which directly improves the asset management or decisions required for Enterprise Resource Planning (ERP). Improved maintenance increases the machine availability and productivity of the organisation.

Figure 1 explains the process of PHM for a system-level prognostic. Data captured by sensors contain a large amount of noise, and hence to study the trend signature from raw data, it must be pre-processed. To study the data, prognostics requires features of this data for analysis. Diagnostics provides information about fault generation and interpretation about the severances. Prognostics provides information about the remaining functional life of the system. RUL predicted by prognostics aids in decision making about maintenance planning.

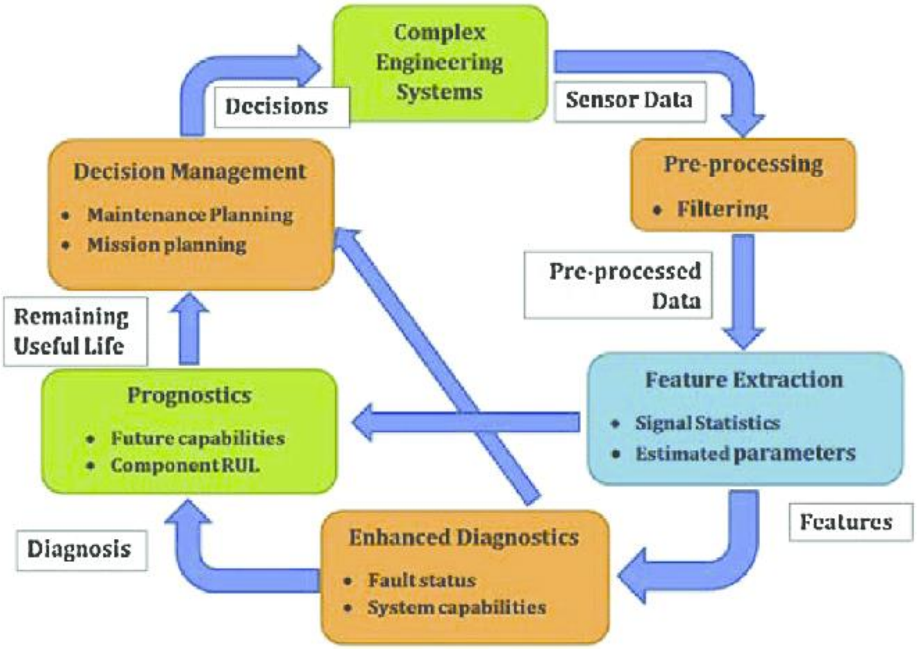

Figure 1. Prognostics process within PHM loop [3]

PHM assistance is essential in large industries such as power generation industries.

Example:

Wind turbines consist of weighty components such as a gearbox, bearings. In case of any breakdown, it is difficult to replace these components, since these wind farms are located at remote locations. Prognostics estimates the RUL of these critical components to prevent the breakdown in the system. Hence implementation of PHM can save organisation downtime and extra maintenance cost.

Industries are applying smart technology to gain the advantages of industry 4.0. Prognostics is one of these sophisticated sciences which can evolve the old maintenance techniques.

As discussed before PHM requires CMD for prediction of RUL. However, in India, there are many small-scale industries which do not use an electronic data collection process. Many industries use human efforts for data documentation on paper, and this makes it challenging to implement prognostics. Lack of proper training to machine operators makes it more challenging to collect the data correctly.

1.2 Prognostics approaches

Prognostics approaches mainly divided into three types, **Error! Reference source not found.**1.2 shows these types using a block diagram. These approaches are based on the input information used by the prognostics approach, i.e. either physical properties or the condition monitoring data or combination of both.

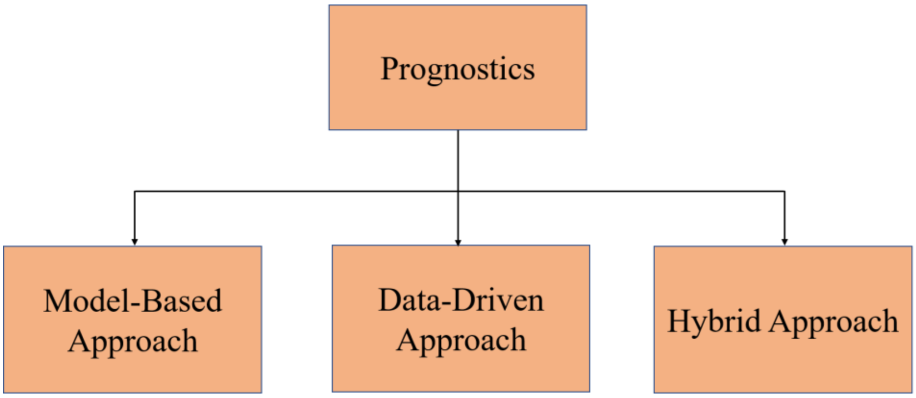

Figure 2. Different Prognostics Approaches [4]

1] Data-Driven Approach

Data-driven approaches use the CMD (Real-time and historical), to predict the RUL of the component. The approach uses the datasets to extract degradation trend of the component to predict RUL of that component. Historical datasets are used by prognostics model to compare the current signature with historical data signature. The process requires a massive amount of data for analysis of features in the data and to assess the health of an asset with more accuracy.

2] Model-Based Approach

Model-based approaches focus on the physical aspects of the system rather than a statistical one for the generation of a model which represents the system behaviour. This approach provides better accuracy than the data-driven approaches in the situation of a simple system. As the complexity of the system increases the number of the component in the system increase, hence modelling a single physical model by understanding the whole system becomes more difficult. Most industrial systems favour the data-driven approach due to higher complexity in the industrial machines.

3] Hybrid Approach

Hybrid approaches user gets the combined properties of the data-driven approach and model-based approach. This approach profits in the way that it gets all the benefits from both the data-driven and model-based approach.

For this project work, significant work in the area of data-driven prognostics has been done. The reasons to select the data-driven approach over a model-based approach is that it is easy to deploy and less expensive compared to the model-based approach.

1.3 Importance of data-driven prognostics

Industries are turning towards the fourth industrial revolution; hence the use of sensing technology for condition monitoring is increasing. Mechanical systems generate a large amount of data while operating. CMD reflects the operating system parameters and performance. Data-driven prognostics establish a model which connects these parameters to the healthiness of the system and estimate the RUL [4].

A considerable amount of data is required by Data-driven prognostics approach about the system in each healthy state, faulty state and transition of healthy to faulty state, which is not easy to obtain. Nevertheless, once the data collection is done, the modelling can be done in a short period. Hence with the right amount of data available data-driven prognostics is easy to deploy.

Several data-driven prognostics models mentioned below require an abundance of data for better accuracy.

Several data-driven models for prognostics:

Independent Increment Process-Based Model

Markovian Process-Based Models

Filtering-Based Models

Regression-Based Model

Proportional Hazard Model

Threshold Regression Model

Several Industries using data-driven prognostics for RUL estimation:

Aerospace industry

Manufacturing Industry

Electronics Industry

Mining Industries

Energy Industry

Physical modelling requires a simple system, whereas, for complex systems, it is impossible to fit a single model which incorporates all characteristics of the system. Modern engineering systems are overwhelmingly complex because of increasing requirements on their functionalities and qualities [5]. Hence data-driven approaches have found superior in such a situation. Unlike the model-based approach, it does not require to know about the physical properties of the system or the representation of parameters. The only requirement in the data-driven approach is that the data must have a specific trend over operating time [6].

Data-driven solutions can provide a piece of unique information about the system and the root causes of fault generation. This information can help to design the component or can help an operator about handling the asset component. Also, data-driven prognostics is inexpensive, accurate and can be quickly developed.

1.4 Challenges in data-driven prognostics

Though data-driven prognostics is easy to deploy, the first step of this approach is to collect the CMD for analysis. This data collection process is cumbersome and time-consuming since the mechanical component takes a long time to reach failure from a healthy state. To collect the CMD from an asset, it must run from a healthy state to failure, but it is not an economical way to do it since the cost of precision-made industrial components is very high.

Mostly in big industries such as power industry maintenance, of critical systems such as a gas turbine or a wind turbine is not an easy task after failure. This downtime of the asset costs the organisation a large amount of cost and power shortage which is harmful for organisation reputation.

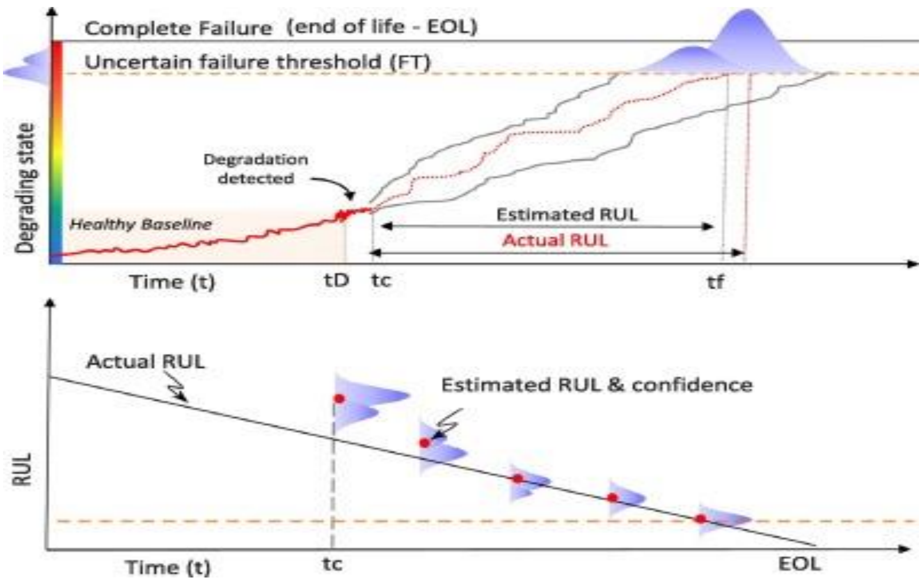

Figure 3. Product life degradation [7]

Figure 3 shows a degradation trend with respect to the operating time. The signature shown in the upper graph shows degradation detection at the current stage. Data-driven prognostics use historical data and predict RUL with a specific confidence. To increase the confidence level of prediction, a large amount of data containing information about the degradation is required.

1.5 Problem Statement

Unavailability of data reduces the data-driven prognostic algorithm's accuracy up to a large extent. This is the reason why many industries cannot apply prognostics models in their organisation. Also, it is challenging to collect this data using conventional methods. Prognostics is an essential part of the fourth industrial revolution. Indian industries are incapable of applying the prognostics due to scarcity of useful prognostics data.

The critical problem of unavailability of industry-grade prognostics data is addressed by this thesis. The problem statement of this thesis is to design and develop such a mechanism which can help to resolve this problem of scarcity of the data using a handful of historical CMD.

A UI must be designed and developed for controlling the hardware and real-time data visualisation of the data generated by the mechanism. Several mechanical components listed in that UI will serve the purpose of selection of component for which data is being generated.

A mechanism has generated which uses a simulation process for new data generation with the help of historical CMD. The simulator can generate prognostics data for several mechanical components. Hence it is named as a GPS.

1.6 The organisation of the thesis

The previous chapter discusses the importance of prognostics for industry 4.0, assisting the maintenance of assets. Chapter 1 also discusses the different approaches of prognostics with the importance of a data-driven approach. Chapter 2 mentions some essential research papers which provides a better insight into the problems and some possible solution to the problem of scarcity of data.

To work in data-driven prognostics, one needs to have basic knowledge about the topic. Chapter 3 provides an overview of data-driven prognostics and identifies the requirement of industry-grade data. Some conventional methods of data collection and their drawbacks are mentioned in the third chapter.

Chapter 4 contains the methodology of this thesis and discusses the generic algorithm for the GPS and hardware needed for its working. Chapter 5 validates the methodology by the results and discussion. Chapter 6 concludes the thesis work and provides information about possible future work for improvisation of process.

Chapter 2. LITERATURE REVIEW

For identification of the problem in the data collection process in prognostics along with the disadvantages due to lack of data, a comprehensive literature survey has been done. This literature survey includes the origin of prognostics and conventional data collection methods to novel methods of data collection.

2.1 Literature Survey

Industrial machines often consist of intricate operating systems and connection between components and subsystems [8]. Systems are required to maintain high-reliability or prone to safety hazards and disastrous consequences [9]. Machine breakdown cost for a single day in a big industry may cost up to 200000 euros [10]. CBM was first introduced in the 1940s. CBM has been seen as an improvisation over preventive maintenance by saving the cost of maintenance [11].

(Peng, Dong, & Zuo, 2010) was one of the first researchers to wrote about the status of prognostics in condition-based maintenance. They also established the relationship between reliability, RUL and the maintenance cost of the asset [12]. (Gillespie, 2015) discusses all the improvisation of CBM [13]. PHM emerged after the use of CBM. One of the leading research centre in the field of PHM, Centre for Intelligent Maintenance Systems, has created more than $855 M of the economic impact on the industry [14].

The twenty-first century is the age of information. Old simple measurements such as oil viscosity or vibration amplitude were used to provide some valuable information about the system. However, the evolution of computers and sensing technology have changed the manufacturing sector, and now data collection methods have improvised using electronic data collection recorded in computers without human interference provides more in-depth insight into the asset condition [15].

Nowadays, industries are talking about advanced ways for maintenance of assets. Industry 4.0 has affected the maintenance approaches for industrial assets. (Ferreiro, Konde, Fernández, & Prado, 2017) discusses the requirements of machining sector for the 4th industrial revolution. They also state that there is a need for new ways to diagnose the fault in assets such as prototype test rigs can be used to understand different failure modes [16].

Prognostics can assist in future maintenance planning, providing the advantages in improving safety, maintainability, reliability and affordability [9]. However, implementing technology such as prognostics in the real industry directly from research face a lot of problems [17].

Prognostics is categorised in three major approaches which are Data-driven prognostics, Model-based prognostics and Hybrid approach. The model-based approach provides better results for simple systems [18]. However, most industries are involved in overwhelmingly complex systems. A simple system cannot fulfil the increasing requirements on their qualities and functionalities [6].

Data-driven methods involve monitoring and analysis of functional product parameters. A data-driven approach for prognostics is recommended when models are not available or when monitoring loads and environmental conditions are not possible [19].

Data-driven prognostics to use an abundant amount of data in deep learning models used in RUL predictions of aerospace systems [20]. NASA finds prognostic technology beneficial for the projects consisting of launch vehicles and spacecraft [21].

A short number of prognostics datasets are unable to train data-driven prognostics effectively. This results in poor accuracy of predicted RUL [22]. Component degradation data measured during service of asset assists prognostics models to analyse the degradation of system functionality. However, collecting degradation data online from industrial systems in practice is expensive and complicated [23].

Collection of CMD till component failure is a time-consuming process. (Skima, Medjaher, & Zerhouni, 2014) use the accelerated life conditions data to predict the RUL of MEMS devices [23] [24].

Primary problems data-driven prognostics technology while implementation in the industry is the unavailability of industry-grade data, change in technologies as well as the design of component in a short period, no communication between actual maintenance practitioner and the researchers. This problem is occurring due to incorrect data collection methods or the absence of data collection in the industry [25].

At the situation of limited datasets availability, model-based approaches also fail due to the unavailability of data for validation of the model. Also, the model-based approach is expensive [12].

(Wang, Yu, Siegel, & Lee, 2008) Discuss a novel methodology for the generation of new datasets which will be more realistic. The author considers the addition of noise into the data for making it more real [22]. A similarity-based data collection is used for the augmentation of data. For this purpose, data having a similar signature and origin are compiled together [26].

2.2 Research Gaps

Several papers mentioned above work on the estimation of RUL of the component or system. A significant problem in the implementation of PHM is the unavailability of the data. Unavailability of data makes data-driven prognostics less accurate. However, to improve this accuracy, some papers used collaborative learning approach or similarity-based approach. Also, manufacturing of testing prototypes was suggested for data collection and diagnostics.

Several papers study the system behaviour in accelerated life conditions. Though this process is timesaving, it may not be an economic one. Many papers talk about the collaborative prognostics but, unavailability of similar assets for collaboration makes it impossible to create a large amount of data for prognostics. Survey noticed the absence of the papers regarding new prognostics data generation from old CMD. It was observed that research is required for the generation of the training mechanism, required for training researchers in proper data collection.

These are some of the primary reasons that industries are not using prognostics. Through this thesis work, a solution is proposed, i.e., GPS, which can fill the gaps found in the literature.

2.3 Research Objectives

1. To develop a mechanism which can resolve the problem of scarcity of quality prognostics data by generating new prognostics with the help of a handful of CMD to generate new data.

2. To generate industry-grade features in the newly generated prognostics datasets without even running the industrial asset.

3. To make the GPS generic so that it can generate data for several mechanical components.

4. To design and develop a User Interface (UI) for GPS web tool for easy access to the GPS from any location and data visualisation.

5. To validate the proposed methodology using a real dataset.

Chapter 3. OVERVIEW OF DATA-DRIVEN APPROACHES

Data-driven prognostics approach uses life data, i.e., CMD of an industrial asset to estimate the RUL of that asset. The procedure consists of pattern recognition using statistical data analysis techniques to make legitimate inferences about the future life of that asset. Model-based approaches focus on the physical aspects of the system rather than a statistical one for the generation of a model which represents the system behaviour. Unlike a physical-based approach, data-driven approaches use information from current CMD data to identify the current state of an asset and use historical CMD to predict the future state of that asset.

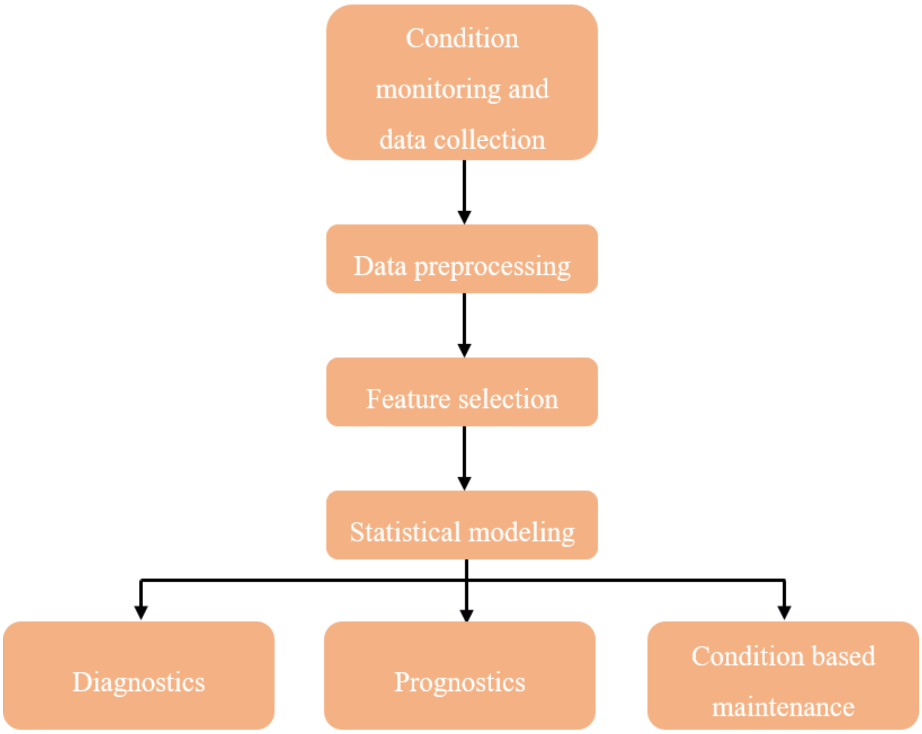

Figure 4. The typical flow of data-driven prognostics systems

In Figure 4, PHM hosts three significant tasks which are diagnostics, prognostics and condition-based maintenance. Diagnostics assists in the identification of the root causes of the fault. This valuable information improves the prognostics and helps to design the system. Prognostics use the processed data for prediction of future failure of the system. This RUL information aids in effective maintenance scheduling by CBM.

To fulfil all these functions, data collection and data pre-processing are crucial steps.

Table 1. Features used for prognostics data

Sr. No.	Feature
1	Mean
2	Root Mean Square
3	Standard deviation
4	Peak value
5	Standard deviation

Data-driven prognostics approach is especially suitable in case of abundance of the Run To Failure (RTF) data. Data-driven prognostics model is required to be trained to achieve the with higher accuracy by machine learning model. Historical RTF data serves the purpose of training the machine learning model used for prediction of RUL.

Lack of knowledge or imperfect measurement causes the unavailability of operational condition or modes of failure, which does not affect the data-driven prognostics [27].

The first step of data-driven prognostics is the condition monitoring data collection. Hence the error in the data collection process can cast the entire prediction of RUL.

3.1 Industrial data for prognostics

Data-driven prognostics approach requires intensive data riches with the information about the state of the system. The system provides a change in the data pattern when it shifts from a healthy state and a degraded state. Various types of industries contain several critical components that are responsible for the failure of the whole system. Such unidentified failures cost the organisation loss of productivity and miss the deadlines for delivery of products or services. For prognostics of system RTF data of these components is required. The primary task is to collect data using the most sophisticated method to find the parametric changes in this data accurately.

Through constant inspection, the observed health indicator is usually referred to as condition monitoring CMD. CMD may directly or indirectly reflect the system health status.

Examples of CMD are the amount of cutting tool wear, chemical viscosity, size of a fatigue crack, vibration amplitude of bearing and the light intensity of the bulb. As mechanical system components degrade with respect to usage, its health deteriorates. This deterioration can be verified using CMD.

Significant types of data used for prognostics:

Waveform type

Vibration, Acoustic emission

Value type

Temperature, Pressure, Humidity

Multidimensional data

Images, X-ray images

Vibration is one of the most significant features used for fault detection in wind turbines. The Root Mean Square (RMS) and peak values of vibration of wind turbines help in estimating the RUL of the turbine and schedule the maintenance activities accordingly [28]. For low-speed bearings, vibration signals are challenging to analyse. That is why acoustic emission is a more critical factor for health assessment of low-speed slew bearings than a vibration [29].

Complex mechanical systems such as rotorcraft drivetrains often receive breakthrough information from vibration and acoustic emission data assisting the prediction of upcoming failure [30]. Electronic companies often use prognostics data for the verification of criticality of electronic components. Health assessment of soldered patches is done using temperature measurement, which aids in monitoring the soundness of electronic components and predicting their RUL [31].

Table 2. Prognostics Data features of several mechanical components

Sr. No.	Component	Common Features
1	Bearing	Vibration, acoustic emission
2	Gear	Vibration, acoustic emission
3	Shaft	Vibration
4	Pump	Vibration, acoustic emission, pressure
5	Cutting tool	Vibration, acoustic emission

3.2 Conventional Data collection Methods

The first step of Data-driven prognostics is the data collection process. The CMD data is collected using different sensors such as accelerometer, acoustic emission sensor, temperature sensor. Data-driven prognostics provide unsatisfactory results in the situation of limited data availability since a low number of datasets are incapable to train prognostics model effectively. The conventional data collection method contains data collected from a healthy state of the component to the failure of that component.

This procedure is time-consuming and expensive. In industries such as power generation industry, it is not possible to run the component to the failure because of the enormous cost associated with the failure of the system. (Example: Gas turbines, aircraft engines and electrical power plants. Failure of such components may cost the organisation economic loss and loss of human life.)

Several customs made accelerated life conditions prognostics simulators are available in the market. Such kits contain actual prototypes of the mechanical systems used in the industry, such as gearbox assembly, bearings assembly. These prognostic simulator setups run in accelerated life conditions. This process reduces the time for the collection of data, but such simulators have some limitations.

Figure 5. Gearbox prognostics simulator

Figure 5 and Figure 6 show the test rigs available for accelerated testing of gearbox and gear train resp.

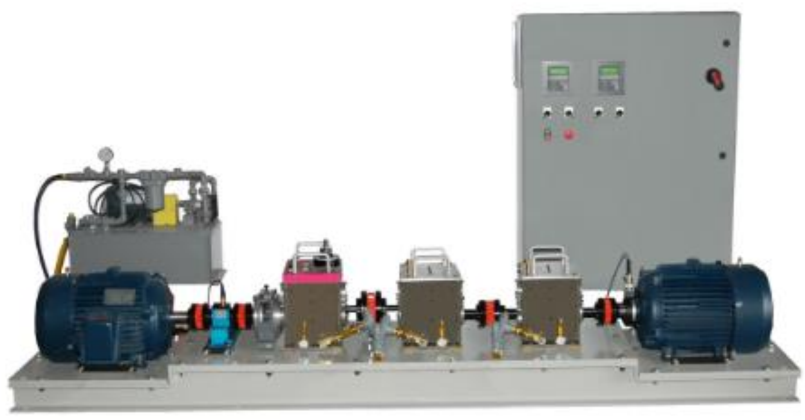

Figure 6. Gear train prognostics simulator

Data collection, with the assistance of various sensors, is the dominant task in the overall process.

Following are the limitations of customs made prognostics simulators:

1] Large capital cost of equipment: Since test rigs are custom made according to the actual system, it may be very expensive.

2] Not generic: These custom-made test rigs cannot generate the data of different components

3] Required planned maintenance: Test rigs require regular maintenance is required for the long life of components.

4] Large size of the setup: Large size of setup makes it difficult to relocate its position.

Chapter 4. METHODOLOGY

The methodology consists of complete detailed information about the data generated using GPS. The algorithm is designed to replicate several features of historical data into new data. Industrial data consist of noise, abrupt changes in degradation and seasonality. Algorithms use historical CMD for training purpose and sense the presence of such features in data and select the features present for new data generation. For a hands-on experience of the data collection process, a hardware setup has been developed which replicates the data generated by the algorithm.

GPS uses a specific process for the generation of datasets. All the process used for data generation is explained below. Every feature used by GPS to generate data, and their requirement is explained in this chapter.

4.1 Prognostics literature survey

Several papers were studied in the literature for identification of the challenges in data-driven prognostics. To resolve one of the problems of data scarcity following work has been done.

4.2 Study of prognostics data

For resolving the data scarcity problem, a better the solution is simulations of prognostics data. For this purpose, a detailed study has been done to understand the various types of data. These datasets are no use in the raw state and hence must be converted into a simple form. Several features were discussed in 3.1 Industrial data for prognostics.

Several datasets of mechanical components such as milling machine, drilling machine were studied for the better insight of industrial prognostics data. Many datasets were in raw format. Several online data repositories were helpful for easy access to the datasets of several mechanical components.

4.2.1 Features of studied data

For the generation of raw prognostics data, it is essential to match the frequency of data matched to the original data with the amplitude. Without this, it is fruitless to generate new data sets. However, significant researchers work on the features of data rather than directly on the raw data. To replicate the datasets with the help of such features was a challenge.

Vibration and acoustic emission are the parameters which can show the degradation of several components such as a cutting tool, gear, bearing. Pressure can be used as a degrading parameter for the pump.

4.3 Development of the algorithm for prognostics data simulator

To develop the algorithms which enable the generation of the new data using the historical datasets, algorithm have to learn from past data. To generate the industry-grade datasets, GPS must incorporate the essential feature of industrial data such as the presence of noise and seasonality. This prognostics data for different assets is time-series data containing different trends, and failure points mathematical modelling is done for generating the data of specific trend and failure points as historical datasets. Steps to generate the simulated data are stated below.

4.3.1 Trend associated with the degradation trajectory

Prognostics data is a time series containing degradation parameter vs time. Degradation occurs in mechanical components while usage until the failure. The deterioration in the component increases with operating time (Example: Crack propagation in gear) as discussed before CMD helps in identification of deterioration. Analysis of this data helps the assessment of the healthiness of the system. Every component fails with the specific trend; hence it is required to generate a methodology which incorporates the sensing of this trend in original datasets.

After sensing the trend in original data sets, the GPS can replicate this trend for new data generation. Since the data-driven approaches depend on the pattern of data, which often has a distinct characteristic near the end of life, it is robust in predicting near-future behaviours, especially toward the end of life [27].

CMD of vibration, acoustic emission, and temperature have some general trends for specific components or systems. These trends may consist of a linear trend or exponential or polynomial trend. The generic algorithm need is to extract the information about the trend in the sample data sets automatically. Failure distribution of mechanical components generally described using Weibull distribution. To decide the life of simulated components, parameters of Weibull distribution must be calculated.

4.3.1.1 Estimation of Weibull distribution parameters.

Reliability engineering uses Weibull distribution as one of the most widely used lifetime distributions. The failure behaviour of mechanical components follows Weibull distribution. The distribution can be in the1, 2 and 3 parameter form. These parameters are shape, scale, and location parameters. This algorithm uses two parameters Weibull distribution for data analysis. Two parameters Weibull distribution holds characteristic life (η), shape factor (β). Characteristic life is a unit time at which system reliability is at 37% while the shape factor provides information about the failure rate.

Table 3. The failure rate in the bathtub curve

Sr. No.	Shape factor magnitude	Failure rate
1	<1	Decreasing
2	1	Constant
3	>1	Increasing

There are several methods for the estimation of Weibull distribution parameters.

i] Maximum likelihood estimation

ii] Least square method

This methodology follows the least square method since the large sample size provides better results in the maximum likelihood method. However, the least-squares estimation method is superior to the maximum likelihood estimation method in the situation of a small sample size of RTF data [32].

The CDF of two-parameter Weibull distribution is given by (i).

$$F(x) = 1 - e^{-\left(\frac{t}{\eta}\right)^{\beta}} \qquad \ldots (i)$$

After calculations;

$$\ln\left[-ln\left(\frac{1}{1-F(t)}\right)\right] = \beta\ln(t) - \beta\ln(\eta) \qquad \ldots (ii)$$

Y = $\ln\left[-ln\left(\frac{1}{1-F(t)}\right)\right]$

X = $\ln(t)$

Equation (ii) is a straight-line equation, with a slope of β and an intercept of β*ln(η). Plot the line to find the distribution parameter.

4.3.1.2 Life of the simulated component

Simulation of time series data requires the time at which component is to fail. This step is responsible for the number of data points in the simulated dataset.

$$TTF = \eta * \left(-\ln\left(Rand(0{,}1)\right)\right)^{\left(\frac{1}{\beta}\right)} \qquad \ldots (iii)$$

Where Rand (0,1) generates, a random number between 0 and 1 with uniform distribution, equation (iii) is the inverse function of the Weibull reliability function used for the generation of unit time for which the simulated component is to run.

4.3.1.3 Start and threshold point of simulated datasets

Manufacturing processes cannot produce the mechanical components with exact dimensional accuracy; this includes the concept of tolerances in the metrology. Hence it is not possible to manufacture 100% quality products always since it is not economical and is expensive. Due to the lack of this accuracy, the health of every component manufactured is different. (Example: Wear of cutting tool made with such precision and accuracy is not precisely zero at zero-hour time). Hence the starting magnitude of the degradation point is different.

After the estimation of the life of the component, calculate the starting and threshold point of the dataset. For this purpose, the application of probability distribution on the starting point and endpoints of historical datasets is done to generate a random number, which represents the starting point of simulation and endpoint of simulation resp.

The generic algorithm generates a random starting point a threshold point using a uniform distribution with maximum and minimum limits collected from starting points of degradation parameter in Sample Datasets (SD). The starting point is declared as Randomly Started Mean (RSM).

$$Random\ start\ mean = random[\min(SD\ start\ points), \max(SD\ start\ points)]$$

$$Threshold = random[\min(SD\ start\ points), \max(SD\ start\ points)]$$

4.3.1.4 Mathematical modelling for trend generation

The prognostics datasets have a start point and threshold and trend connecting both points. This trend extraction process requires the analysis of SD. The algorithm analyses the trend in sample data sets. Equation (iv) shows the mathematical multiplier to generate the degradation parameter values at each Time Stamp (TS). It is essential to analyse the trend of historical datasets by the equation of the best fit line.

$$X = \frac{[Threshold - RSM]}{(TTF)^n} \ldots (iv)$$

Where 'n' is the power according to the historical datasets, (Example: If SD has linear data points, the value of n is 1, If SD has a^{th} polynomial order equation, then the simulation generates X value considering n = a).

4.3.1.5 Change in past Time Stamp data

After the calculation of the starting, threshold point of data, the change in the degradation parameter after each TS is calculated until it reaches the threshold at TTF. Change In Mean (CIM) generates the time series best fit data, which follows the trend as SD follows.

$$CIM = RSM + X * [Instantanious\ time]^{n} \quad \dots (v)$$

4.3.2 Incorporation of Noise in the data

Industrial assets operated under very harsh environmental conditions; hence industrial environment affects the performance of the assets. In data, collection process sensors sense the readings from assets with the irregularities caused by these irremovable conditions.

Since noise contaminates industrial data, it produces errors in the data collection, storage, and analysis. The presence of noise hinders the processing capability of a machine learning algorithm and reduces its predictive performance and increase its training time. There is always a 5 % chance of error in the data in a controlled environment. To prevent errors in the prediction due to noise, researchers often use the data cleaning process before application [33].

Since the noise is an inseparable part of the industrial data, to generate realistic datasets, it is required to inject the noise in simulated datasets. This injection of noise has a random nature. The inverse function is used to incorporate noise in the simulations. Random numbers generally follow the normal distribution. Hence the generation of irregularity in the data is injected by the inverse normal function.

Required parameter for the generation of noise is mean and standers deviation of specific TS. CIM is the trend line of the simulated dataset; hence it is the mean value for inverse function. Calculate the maximum percentage error of actual value to the trend line in sample dataset to calculate the standers deviation values.

$$Noise = normalinverse[Rrand(0,1), CIM, CIM * (A\%)] \quad ...(vi)$$

Where A is the maximum percentage deviation of sample datasets to respective trendline mean.

4.3.3 Incorporation of abrupt jumps in the data

According to the change in operating conditions, degradation trajectories usually exhibit several features due to sudden shocks and physical mutation. This sudden changes in degradation often appear at the change in the system (Example: While the change in the phase of the workpiece being machined), which makes the estimated value of RUL unreliable. This is the reason that GPS should be able to generate jumps in degradation data.

The sudden jumps in the degradation data required to be analysed before the data generation. If a cutting tool working tip breaks the vibration reading increases drastically. This jump in the data is random but generally has some probability of occurrence. This jump probability is calculated from the historical datasets. This jump probability is used to create the jumps in simulated data.

For this purpose, a jump probability has been set for the generation of drastic jumps in data. The slope of the SD is useful for calculating the jump probability. Calculate the slope of data points at each TS. A significant value of the slope is considered a jump. Calculate the percentage jumps in the data points. This percentage of jumps is the jump probability for the generation of datasets. Add this new jump to the data points with added noise respective that TS.

4.3.4 Seasonality addition in the data

Due to the repetitive nature of several operations of industrial assets, it encounters similar operating conditions. Such situations are responsible for the generation of a similar trend of data. Example: Milling cutter produces similar data at the identical position of each number of cuts, which generates seasonal variation in the condition monitoring data.

A study performed by (Davey & Flores, 1993) proposes a method for finding the presence of seasonal variation using statistical tests. They used statistical correlation analysis for identification of seasonality in the data with higher confidence [34].

Several ways to identify the presence of seasonality in a time series:

1. Knowledge of the product (such as number of phases in the workpiece),
2. Statistical analysis of the data.

The seasonality in time series data affects short term forecasting up to a large extent. It is also useful in the short-term policy decisions of the organisation [35].

The seasonality multiplier is calculated from historical datasets to implement Seasonality. These multipliers can be multiplied to the simulated datasets to get more realistic data with seasonal variations.

4.4 The architecture of the proposed prognostic data generation model

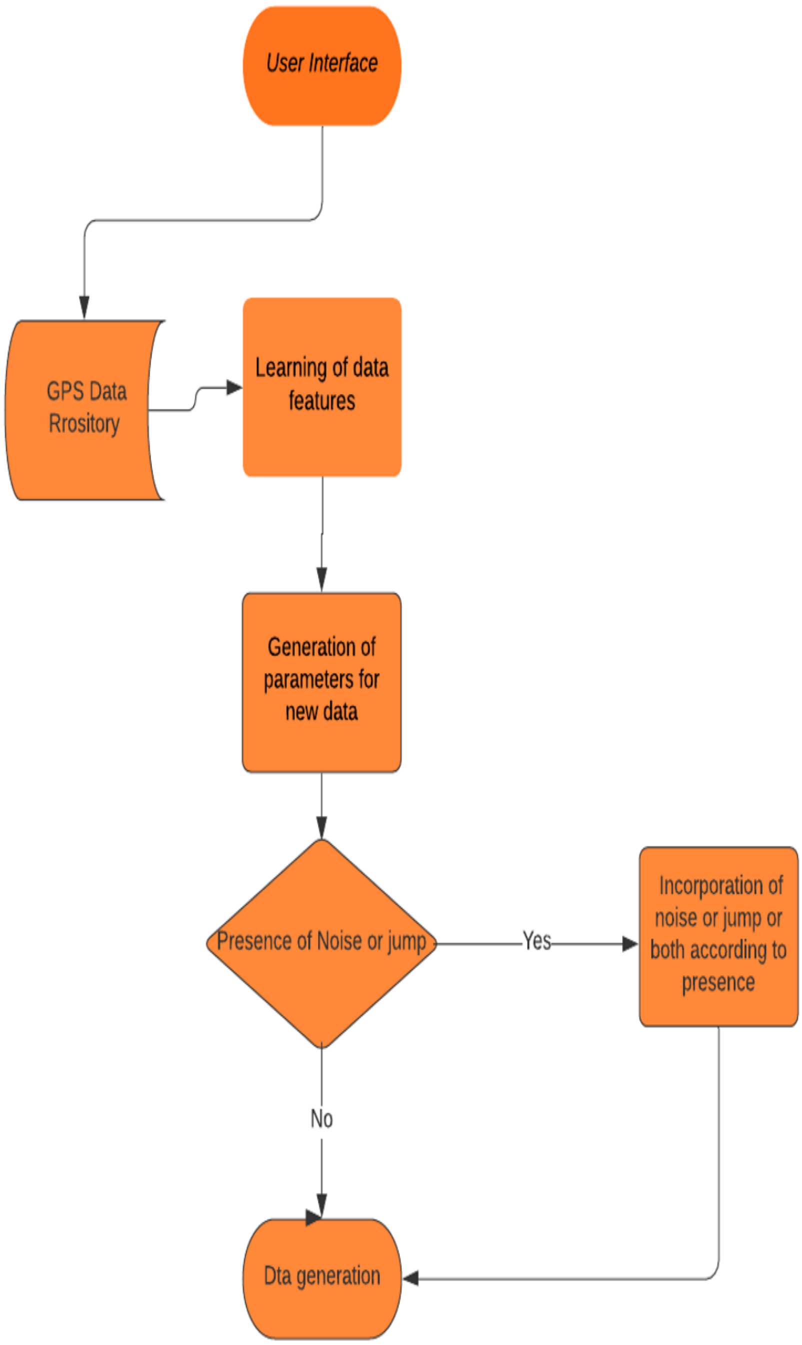

Figure 7. The architecture of the proposed prognostic data generation model

4.5 Mechanism for the actual generation

A mechanism is required which can be controlled using the commands of a programmed algorithm. A controller can use these commands to control this mechanism, and with the help of the measurable output data of that mechanism, replication of simulated data can be done. For this purpose, a parameter is needed, which can be sensed using sensor and data can be stored. Vibration is one of the parameters which can be generated using a mechanical system. Vibration can be measured with the help of accelerometer.

Systems able to create controlled vibration signals:

- Electric motors
- IC engine

After calculating the vibration RMS of DC motor at different speeds, it shows the consistency in the RMS at the respective speed, which makes the DC motor reliable for this purpose.

Advantages of the electric motor over the engine:

1. Economic
2. Clean energy
3. Compact in size

The generation of datasets can be done on the simulation algorithm. However, for training purpose of researchers, hands-on experience is required for the data collection process in prognostics. This is the reason why GPS is the combination of both the software and hardware. Software part already has been discussed in the points above—a hardware mechanism required for the replication of the data generated by simulations. A detailed explanation of the hardware setup of the GPS is done below.

Pieces of equipment required for the GPS:

- A Direct Current (DC) motor
- Arduino UNO board

- Motor driver
- Uniaxial accelerometer
- Piezoelectric coupler
- Data acquisition card

4.5.1 A DC motor

A plain DC motor is used for the experiment since it is compact and easy to supply power to the motor.

Figure 8. DC motor

DC motor is extensively used industrial element for speed control and load characteristics. It is easily controllable and provides precise output; hence it is widely used for the commercial purpose [36].

Table 4. Specifications of DC motor

Sr. No.	Property	Value
1	Voltage range	6-48 volts
2	RPM range	2400-12000 RPM
3	Max load current	1000ma
4	No of Poles	2

A small DC motor is used for secure handling of GPS and to keep system compact. To control this DC motor algorithm will command the controller which control the speed of the motor.

4.5.2 Arduino UNO

Arduino UNO is a microcontroller-based electronic board which can be programmed using Arduino Integrated Development Environment (IDE). User can use Arduino A type port to connect the board with a computer to control the microcontroller of Arduino using programs in IDE. A motor can be controlled using a motor driver circuit with the help of analogue and digital output of Arduino.

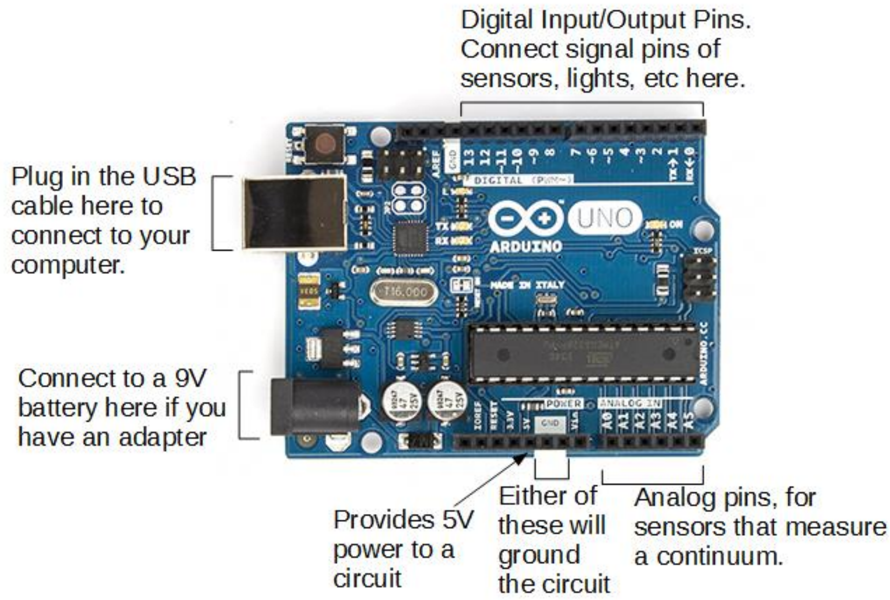

Figure 9. Arduino Circuit Board

Following variables can control the speed of the DC motor:

1. Voltage
2. Flux
3. Resistance

Controlling the input voltage supply of the DC motor can be an easy way to control the speed of the DC motor, as Voltage increases the speed of the motor increases. Standard voltage control can cause a lot of power loss on the control system of applications, so the PWM method is widely used in the DC motor speed control application [36].

The basic principle of PWM is to switch power on and off at a specific frequency to maintain the duty cycle at the required percentage. The terminology of the Duty Cycle is the ratio of 'ON' time to the cycle time. Duty cycle is specified in the percentage format. Higher the duty cycle higher the power.

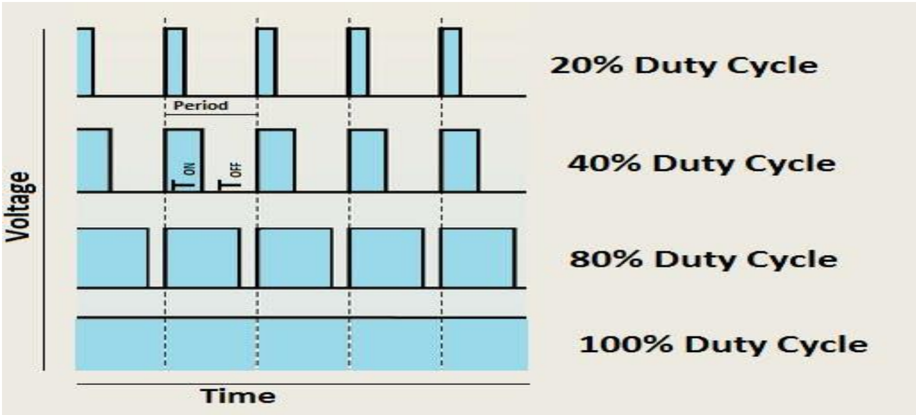

Figure 10. PWM duty cycle

4.5.3 Motor Driver (l298n)

A motor driver is required for adequate control the motor speed. The circuit will allow the user to quickly and independently control DC motor up to 2A in both directions. This motor driver l298n uses the Arduino PWM output as the input. The detailed connections are shown in the figure in appendix C.

Table 5. Specifications of motor driver l298n

Sr. No.	Property	Magnitude
1	Input Voltage	3.2V - 40V DC
2	Peak output current	2 A
3	Channels	2
4	Operating current range	0 – 36 mA
5	Storage temperature	25 – 130 °C

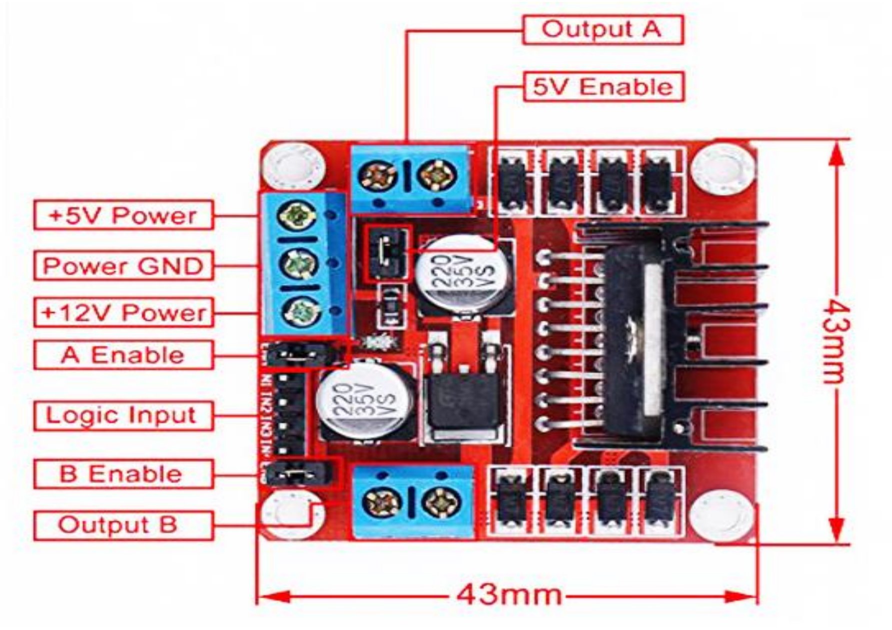

Figure 11. Motor driver connections L298N

4.5.4 Piezoelectric accelerometer

A uniaxial piezoelectric accelerometer is used for sensing the vibrations signals of DC motor.

Figure 12. Accelerometer with a magnetic mount

Manufacturer: Connection Technology Centre, Inc.

Model: AC102

Table 6. Specifications of the uniaxial accelerometer

Sr. No.	Property	Magnitude
1	Sensitivity	100 mv/g
2	Frequency response	30*900000 CPM
3	Dynamic Range	± 50 g
4	Power requirement	18 – 30 Vdc
5	Temperature range	-58 to 250° F

4.5.5 TEDS Piezoelectron Coupler

Small DC motor can not produce high amplitude vibration signals. The magnitude of vibration signals of the DC motor and the magnitude of the actual data has a large gap. To convert the magnitude of data generated by DC motor vibration up to the magnitude of actual data, the motor vibration data must be multiplied with a multiplier. For this purpose, a piezoelectric coupler works as an amplifier which amplifies the vibration signals of DC motor.

Figure 13. Kistler TEDS Piezotron Coupler

4.5.6 Data acquisition card (DAQ)

A data acquisition system is used to store the signals from an accelerometer into the computer system.

- Model: cDAQ9188XT
- Max. sampling range: 100Hz to 50KHz

Figure 14. Data acquisition system

4.5.7 Experimental Setup of GPS

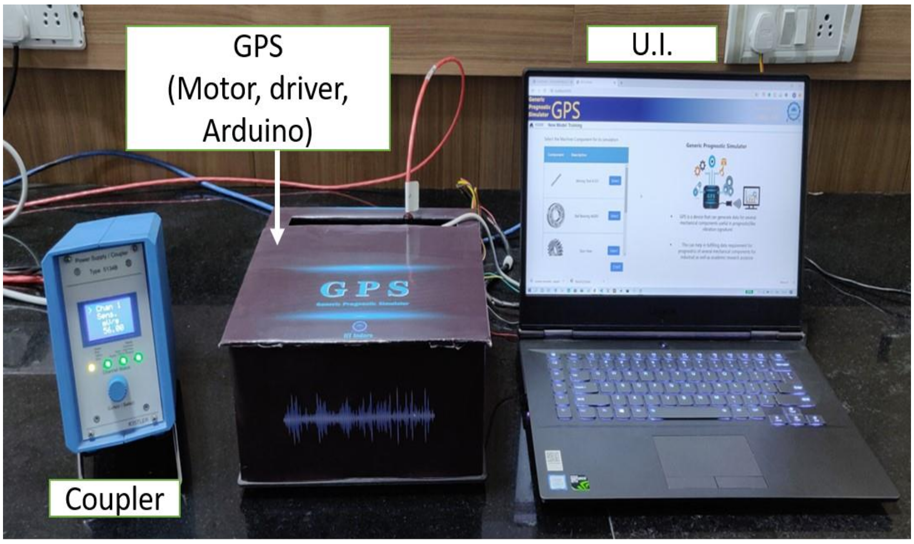

Figure 15. User interface and hardware setup of GPS

Experimental setup of GPS is shown in which consist of the UI of GPS. The housing is shown in the named as GPS consist of DC motor, motor driver l298n, Arduino UNO and uniaxial accelerometer attached to the base on which motor is mounted.

4.6 Validation of DC motor vibration readings

DC motor produces specific vibration levels at specific Rotation Per Minute (RPM). For validating this, an experiment was performed in which the speed of DC motor was increased from 0% to 100%. This can be done by changing the duty cycle by increasing Analogue Write Value (AWV) from 0 to 255 as discussed in 4.5.2 Arduino UNO

4.6.1 Generation of the databank

To generate similar values to simulated datasets, the motor needs to run on specific speeds so that the accelerometer catches the vibration signals at that respective speeds. Vibration signals for different speeds of the motor are required to be collected as a databank. It is not possible to generate the data with the same frequency as other industrial assets generate. Hence most of the researchers work on the vibration RMS signals.

One common expectation of PHM is its capability to transform the raw CMD into actionable information, to facilitate easy maintenance decision making [37]**.** For identification of the variation in the two-vibration signal, statistical time-domain features can be used. These features are mean, RMS, Standard deviation and variance [38].

It is easy to calculate the vibration RMS values at each AWV where AWV controls the PWM results in control of the speed of the motor and respective vibration. AWV range from 0 to 255. As AWV increases the duty cycle of PWM increases which is 0% at 0 AWV and 100% at 255 AWV respectively.

As the figure below shows, after 165 AWV, the change in RPM of DC motor is negligible. This affects the vibration signal generated by the motor. After 165 AWV the vibration stabilizes, hence this 165 AWV is a threshold after which stable vibration signals are produced by a motor which is no use for the generation of degradation signature of increasing trajectories of fault.

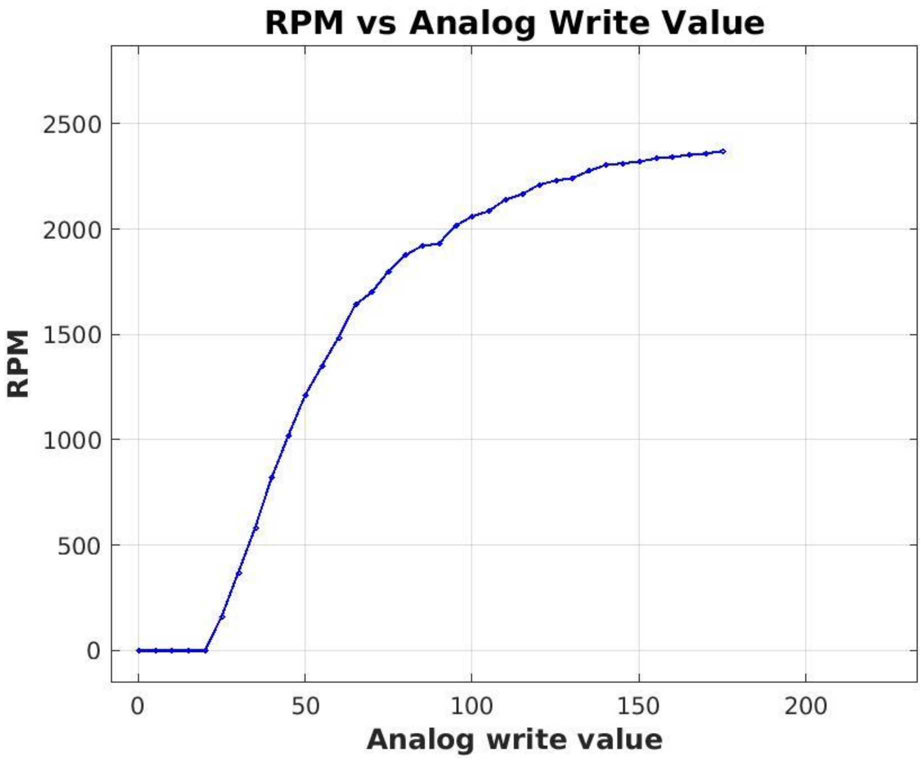

Figure 16. Motor change RPM with an increase in AWV

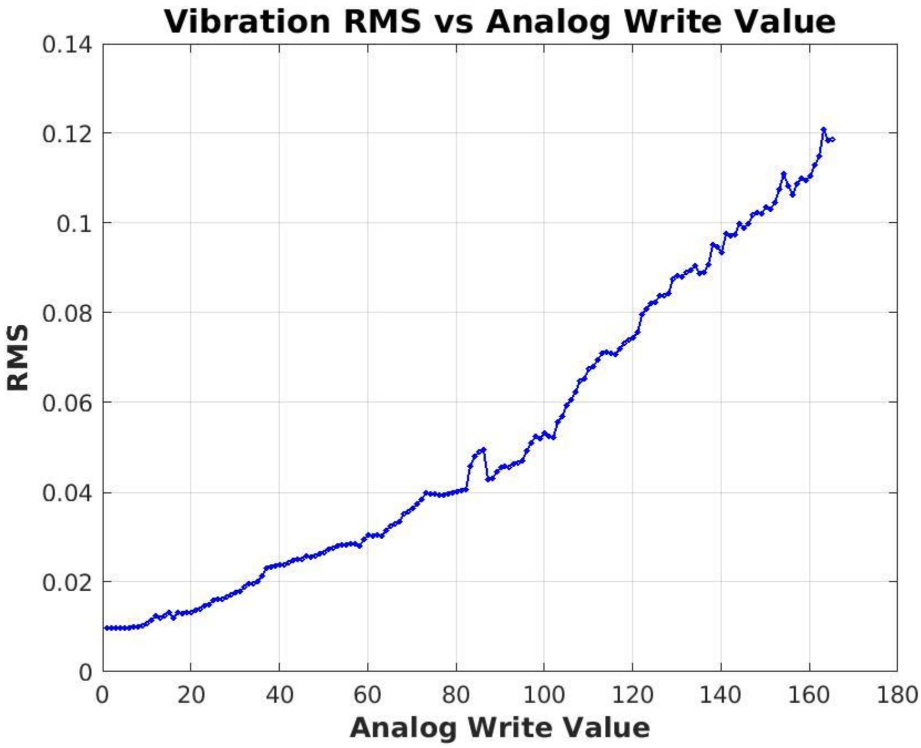

Figure 17. Change in motor vibration RMS vs AWV

This change in vibration with respect to speed helps to generate prognostics data trajectories.

4.7 Validation of simulated data and motor generated data

For validation of this data, a dataset showing the degradation of a mechanical component is required. After studying the data and pre-processing it before applying the algorithm on it, new data sets can be generated. According to the simulated data signature, Arduino sends the signals to the motor driver to run the motor on specific speeds. Thus, the motor rotates on specific RPM vibration generated by motor will match to vibration of specific PRM in the databank.

This vibration RMS helps in the generation of actual data with the exact trend of simulated data. To collect this vibration from the motor accelerometer is mounted on the base plate of motor mounting. Since the motor cannot vibrate at the amplitude of vibration on which industrial machine vibrate. The vibration trend generated by the motor is amplified by the coupler. After amplification data, it is stored in the computer by the DAQ system. The similarity between actual data and simulated data are shown in Chapter 5.

4.8 Generation of Data-repository

To study different data sets, it is a very crucial step to collect the actual datasets form authentic online data repositories. The new datasets collected from online data repositories or collected from actual industry must be collected in a system-based data repository. This collective data repository can act as an independent data repository for GPS.

After simulation of new data sets using GPS with the help of datasets in a data repository, these new datasets can be added to the respective component data in the repository. More the components data in the repository more utilisation of GPS can be done.

4.9 Concept of a generic data simulator

The algorithm for simulation of prognostics data mentioned in 4.3 Development of the algorithm for prognostics data simulator can be used to generate data of several mechanical components such as milling machine, drilling machine. However, the experimental setup developed so far can only provide hands-on experience of vibration data collection. For data collection of other data such as temperature or pressure, hands-on experience of data collection is not possible at this moment.

To generate the prognostics data using GPS, one has to follow all the steps mentioned above in the methodology.

4.10 Generation of new datasets using GPS

Any number of datasets can be produced using GPS and the dedicated data repository. A large number of datasets may take few hours to generate the data, but it is very less time than actual data collection on assets.

4.11 Design and development of UI for GPS

For smooth operation of GPS, a user-friendly UI has been designed and developed. For designing the UI, some study was done to know the basic requirements of UI.

UI is an integral part of software or hardware or hybrid system.

An ideal UI must content:

- Easy operation
- Quick in response
- Simple yet multifunctional design
- Effective handling of operational errors.

UI developed for GPS consist of simple design. It consists of a list of components for which GPS can generate the data. As the learning of a new dataset is complete that dataset gets added into this list. User can choose the component in the list for data generation, and in a few seconds, the user can visualize the real-time data generated by GPS. After completion of data, simulation user can save that data to his system.

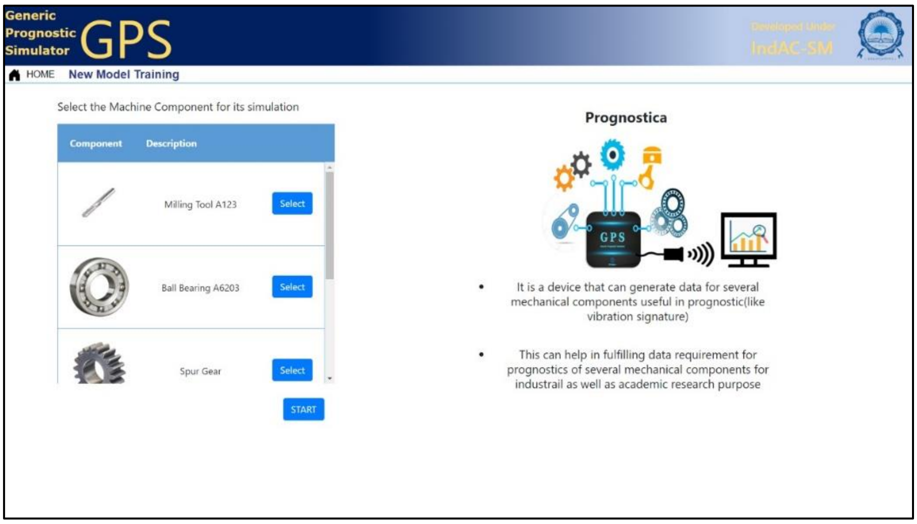

Figure 18. Home page of GPS UI

Any user can access this UI webtool using the internet. A provision of new model generation is provided in UI. New model generation helps a new user to upload his data and develop a new model for new data generation using GPS.

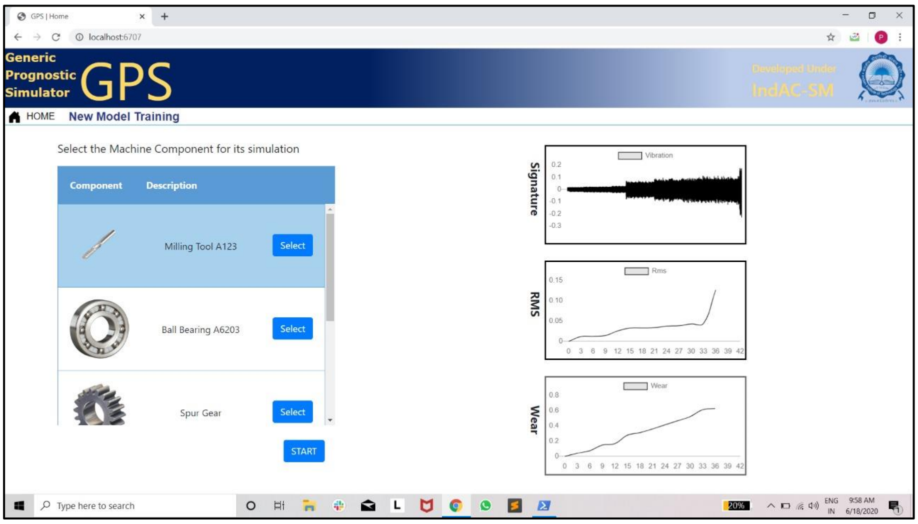

Figure 19. Data visualisation using GPS webtool UI

4.12 Validation of complete GPS by available data by the proposed methodology

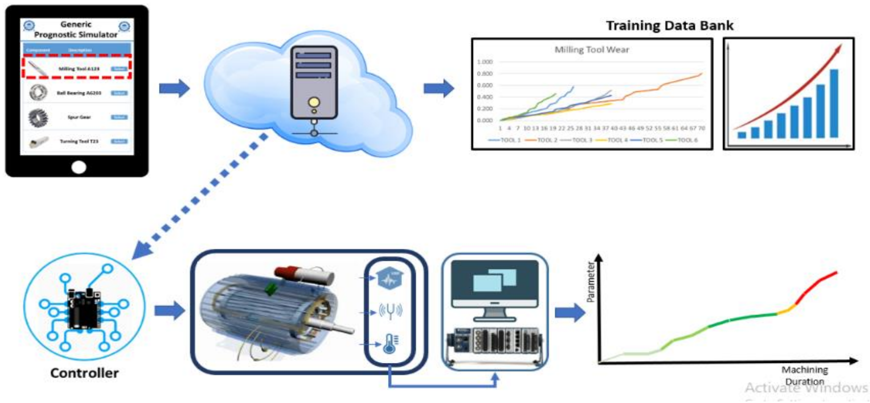

Figure 20. Complete Methodology of GPS

The validation of completer GPS, i.e., backend programmed algorithm, data generation using hardware setup and front-end UI/UX, is done. The data used for validation is shown in APPENDIX C. Also, results have been discussed in the next chapter.

Chapter 5. RESULTS AND DISCUSSION

This chapter explains the details of the data obtained from GPS. As discussed in the previous chapter, every step is followed to get these results.

5.1 CMD Data for validation of the methodology

For validation of proposed methodology data of a CNC milling cutter (Wear and Vibration RMS) was used. EMCO MILL E350 CNC three-axis high-speed vertical milling machine is utilised as the testbed. Normal degradation of end milling cutting tool is carried out to study the degradation behaviour of the cutting tool. A high-speed steel 6 mm milling tool is chosen for analysis. The workpiece used is of mild steel of dimension 165mm x 100mm.

Six end milling tools run to failure data was generated for the following operating condition shown in the table. During the machining, force, vibration and acoustic signal were monitored continuously during every cut of machining the mild steel plate for the length of cut equal to 1320.

Operating conditions:

feed=250mm, speed=1300rpm, depth of cut=0.35

Table 7. Milling cutter life data

Cutter	Time (No. of Cuts)	Failure Modes
1	14	Breakage
2	15	Breakage
3	18	Breakage
4	21	Breakage
5	30	Worn Out
6	31	Worn Out

The graphical representation of milling cutter wear and vibration RMS data is illustrated in Figure 22 and 23, respectively. Each trajectory represents a cutting tool, and the readings were taken after each machining cut.

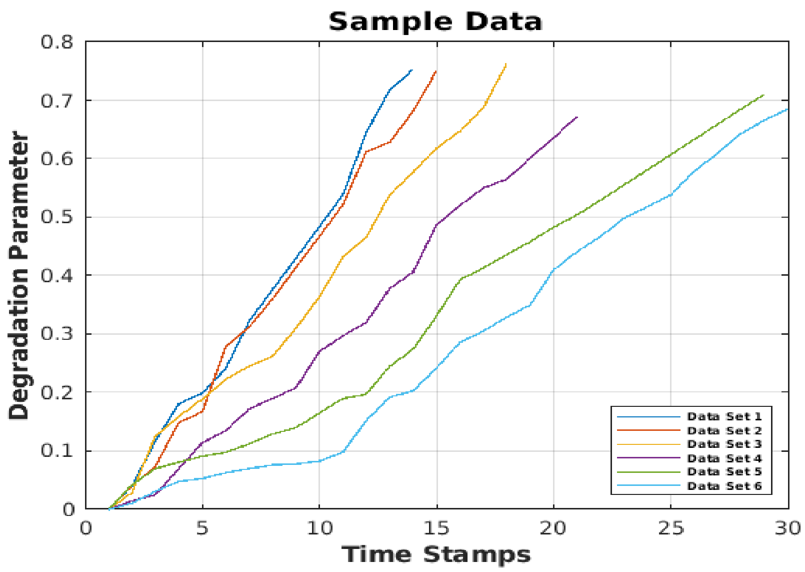

Figure 21. Milling cutter wear data

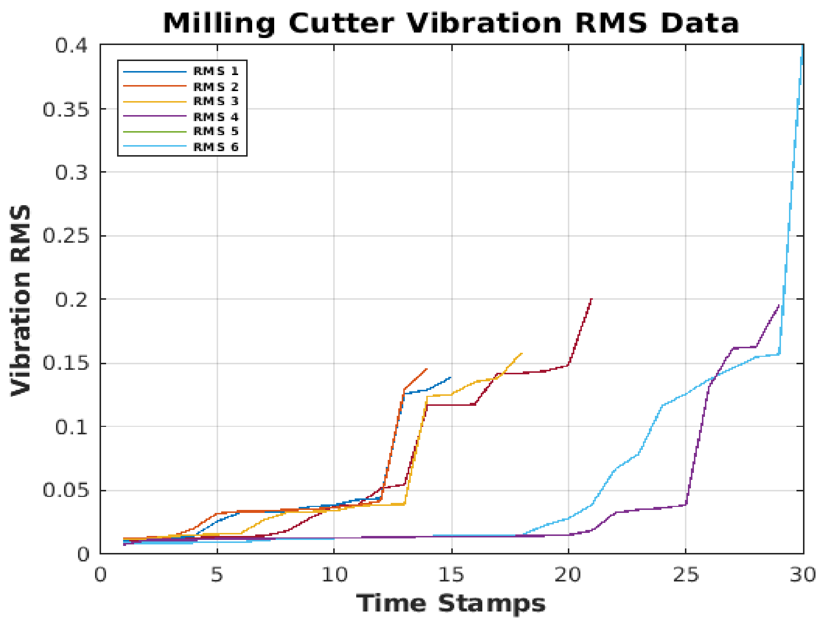

Figure 22. Milling cutter vibration RMS data.

Table 8 Sample wear data of milling cutter

Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5	Dataset 6
0.000	0.000	0.000	0.000	0.000	0.000
0.011	0.042	0.015	0.029	0.041	0.040
0.031	0.070	0.025	0.126	0.074	0.117
0.048	0.081	0.070	0.159	0.149	0.181
0.053	0.091	0.114	0.189	0.167	0.199
0.063	0.098	0.135	0.223	0.278	0.241
0.070	0.112	0.172	0.245	0.312	0.322
0.076	0.129	0.190	0.262	0.360	0.376
0.078	0.140	0.208	0.310	0.414	0.430
0.083	0.165	0.270	0.364	0.467	0.484
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

Table 9. Sample vibration RMS data of milling cutter

Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5	Dataset 6
0.012872	0.011317	0.012058	0.010966	0.007616	0.009141
0.013328	0.011826	0.012171	0.01169	0.01114	0.009141
0.013533	0.012494	0.014541	0.011914	0.011267	0.009207
0.020338	0.014428	0.0153	0.013185	0.011309	0.009207
0.032046	0.025749	0.015944	0.013561	0.012257	0.00953
0.033739	0.03292	0.016397	0.014032	0.012392	0.00953
0.0339	0.03311	0.026946	0.014611	0.012562	0.011222
0.034986	0.033973	0.032911	0.018442	0.012708	0.012248
0.035176	0.037579	0.033481	0.028642	0.012849	0.012409
0.036783	0.038664	0.034165	0.03794	0.012966	0.012619
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

Complete data is shown in APPENDIX C

5.2 Results and discussion

As mentioned above a milling cutter dataset are collected by conventional condition monitoring procedure. It is tough to generate the exact signature from raw data since it is difficult to match the frequency. The vibration data is converted in RMS. Each value represents a milling cut**.** Every mechanical asset shows some significant change or characteristic before the failure, which is an indicator for the remaining life and the severeness of failure. This dataset shows a sudden increase in the vibration RMS before one or two cuts of the failure.

Results show that there are two failure modes which are worn out and breakage. Due to the chipping of tool toll fails at an early stage. The generic algorithm has considered sample dataset information for the generation of new datasets.

Also, the industry-grade features such as jump can be seen in the vibration data.

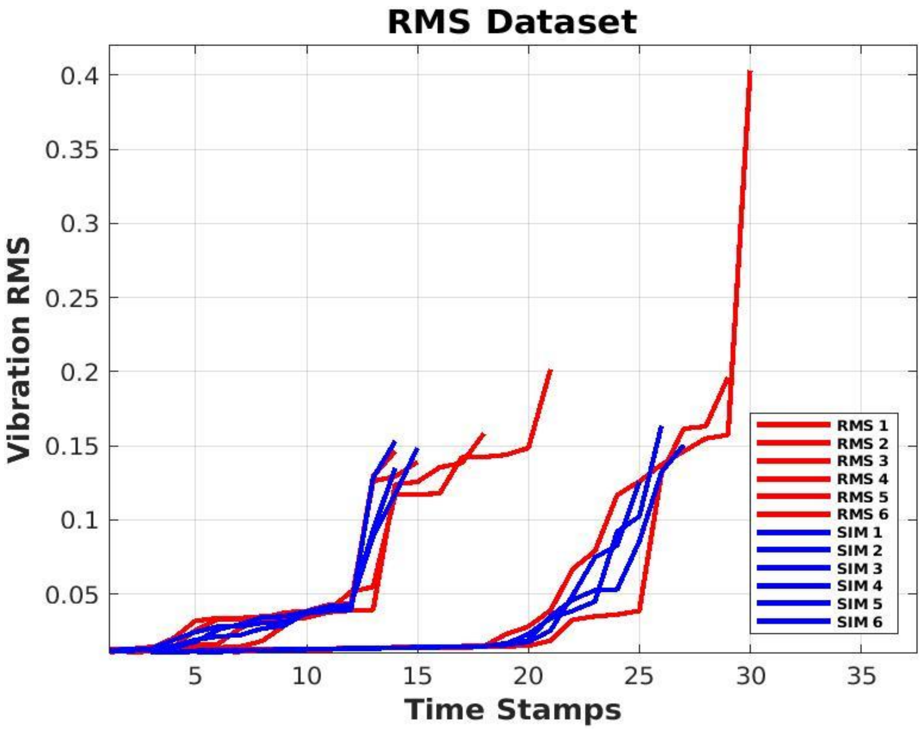

Figure 23. Vibration RMS data generated using GPS

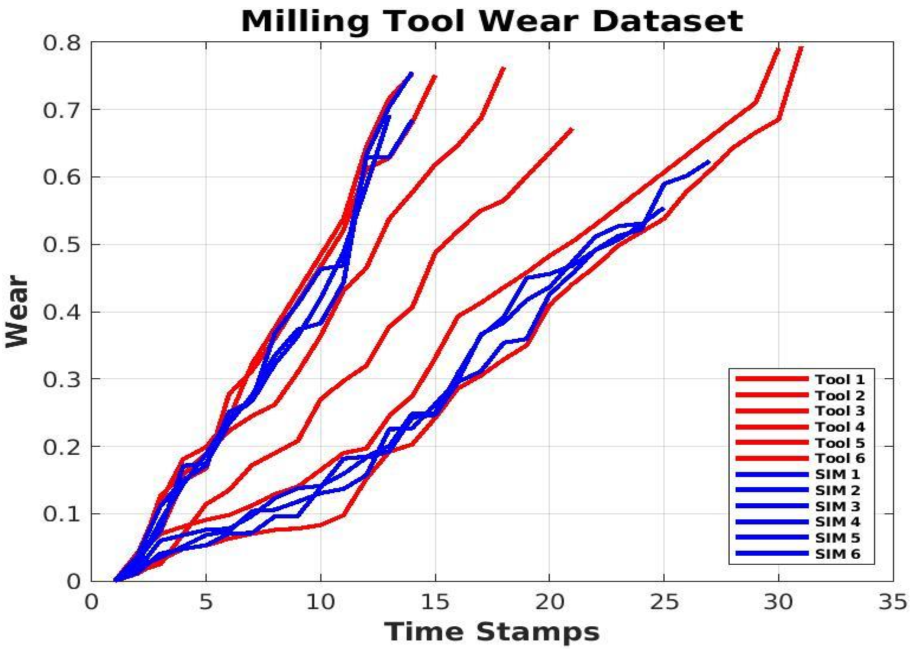

Figure 24. Tool Wear data generated using GPS

It is crucial to find the similarity and error between both the simulated and motor generated data. For both the data of vibration RMS and Wear a single dataset has been generated by simulation and then by motor and similarity and error calculation has been done. The selected wear and vibration data for this calculation is shown in Figure 26 and 27.

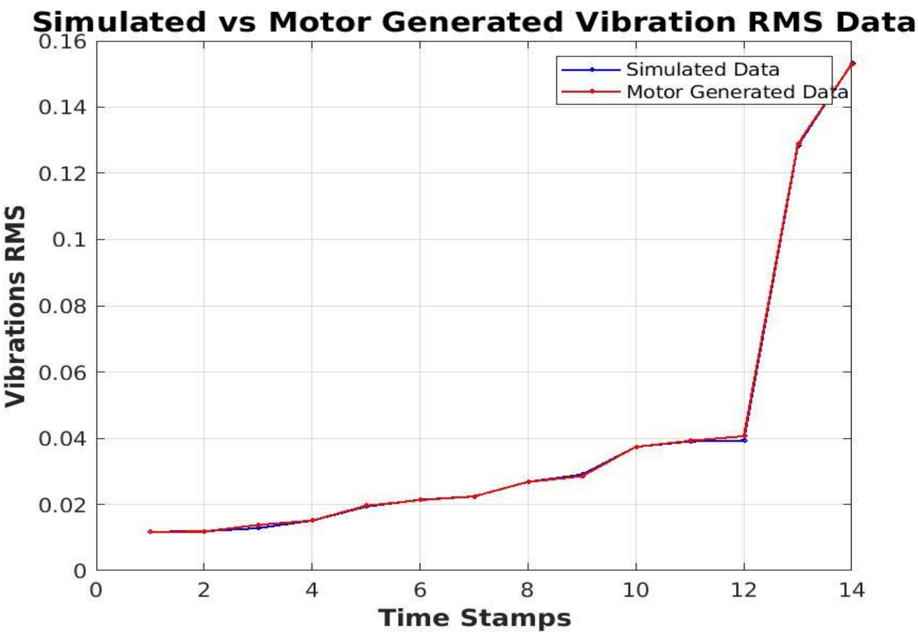

Figure 25. Simulated and respective Motor generated RMS data

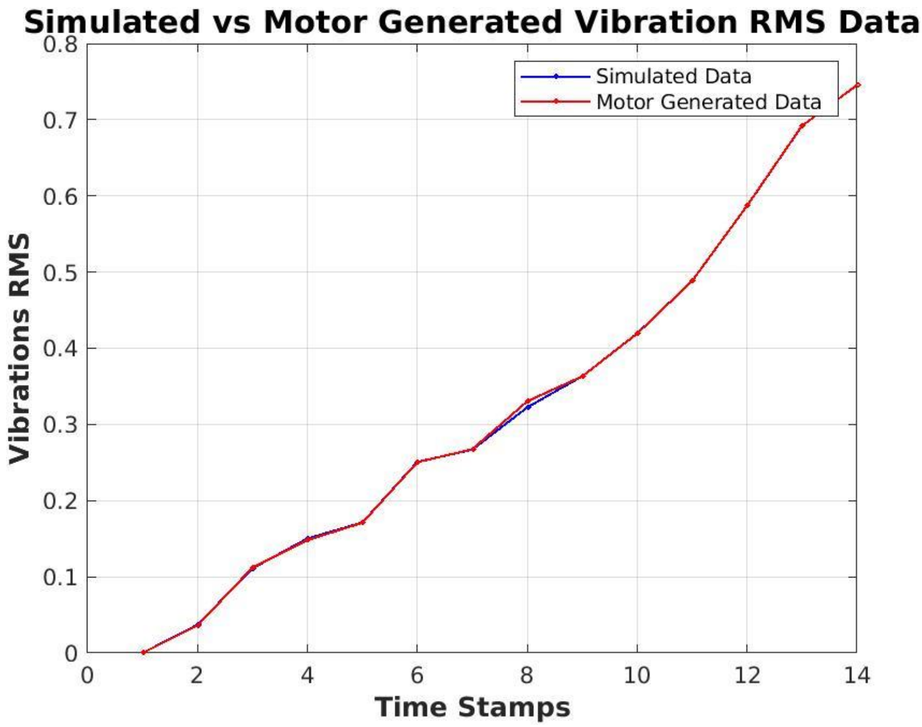

Figure 26. Simulated and respective Motor generated wear data

Calculation of cosine similarity index:

To find the similarity between the results obtained from simulations and the motor generated data, a similarity index must be defined. This similarity can be quantified as the cosine of the angle between vectors, that is, the so-called cosine similarity. Cosine similarity is one of the most popular similarity measures applied for clustering [39].

Given two vectors datasets A and B their cosine similarity is

$$Cosine\ index(A,B) = \cos(\theta) = \frac{A * B}{\| A \| * \| B \|}$$

$$Cosine\ index(A,B) = \frac{\sum_{i=1}^{n} AiBi}{\sqrt{\sum_{i=1}^{n}(Ai)^2} * \sqrt{\sum_{i=1}^{n}(Bi)^2}}$$

-Where A and B are n-dimensional arrays.

For calculation of cosine similarity index between data generated by simulation and by actual motor data, A single dataset has been generated. The similarity index between both simulated data and motor generated data is shown, below the graphs of each wear data and vibration RMS data.

Table 10. Cosine similarity index for simulated and hardware generated data

Sr. No.	Data	Cosine Similarity index
1	Vibration RMS	0.999968
2	Wear	0.999984

RMSE calculation:

For calculating the simulated data and the data generated using the hardware setup. Root Mean Square Error (RMSE) parameter is used for quantification of the error between simulated and motor generated data

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{[(Simulated\ data)i - (Motor\ generated\ data)i]^{\wedge}2}{n}}$$

Table 11. RMSE for simulated and hardware generated data

Sr. No.	Data	RMSE
1	Vibration RMS	0.000489
2	Wear	0.00267

Several prognostics models are available in the market. These prognostics models work on some assumptions. On of such assumption is the random failure of the industrial asset. However, a study of the data induced that the asset component failure is not random. It is a pattern distributed failure, but unavailability of a simulator like GPS creates a problem in prognostics models. Now the GPS has been developed to resolve this problem; hence it can provide a more realistic effect for the prognostics model.

To design and develop such a mechanism for data generation is a novel approach in the research area. This complete process is consisting of an asset component degrading with time, data collection and learning by the algorithm to generate simulated datasets. After simulating to provide a hands-on experience for training of prognostics for the different asset. This process can not be created in the academic background because of the lack of actual assets.

(Example: To generate a prognostics model for gas turbine academic researchers do not have enough knowledge for data collection for gas turbine since the absence of gas turbine.) However, for generating a model for prognostics of the gas turbine, one does not need a gas turbine. Using a GPS with sample gas turbine degradation data, one can train for start to end process.

CHAPTER 6. CONCLUSION

As prognostics is an essential part of industry 4.0. There are several problems in the implementation of prognostics which are already discussed in the literature review. This thesis tries to solve one of the problems associated with the implementation of prognostics. The data collection, to manage the data is not a simple job. Hence among these problems, this thesis work tries to resolve data scarcity problems.

While working on this has been seen the importance of data-driven prognostics approaches. This shows that data is actually the soul of prognostics. Useful data affect the whole process of RUL estimation. Without data, it is impossible to do the prognostics. If data is not of good quality accuracy of prognostics model will be reduced. Hence data is inseparable pat of prognostics. Hence, we thought about the possible ways to resolve this data scarcity problem.

Then the idea of a data simulator came into existence. So, this thesis talks about the complete process for making such simulator. This thesis tries to incorporate several industrial features into the simulated data to make it more realistic. For these purposes, several factors, such as a jump in the data, seasonality in the data. For this purpose, the study of several component data was done. Then the development of a mechanism using a DC motor was developed. Using the steps in methodology, the generation of data was done. The validation of generated data was done. For this purpose, cosine similarity and RMSE was calculated, which were up to the mark.

In this process, our experiment of, use of DC motor to generate such type of data was successful. Now because of GPS, new avenues are opening.

A requirement of a mechanism which resolves data scarcity for prognostics was the primary job of this thesis. This work provides new avenues to the research in prognostics.

CHAPTER 7 FUTURE SCOPE

To fulfil everything mentioned above in the industry, some basic things must be done before the implementation of GPS. Use of federated or collaborative learning can be used for GPS. An independent cyber-physical system can be developed for the communication purpose of GPS with another asset or GPS.

While talking about academic research, it can be done that this GPS can take the decisions and perform the machine to machine communication. Star to end the process of academic research must be done. This device can be the solution of clubbing this process in one device. Being a generic prognostics simulator, only vibration-based data can provide hands-on experience, but similar parameters can be incorporated by GPS such as temperature, pressure

Hence new mechanisms must be found to generate measurable temperature or pressure generation, which can make this GPS a full proof prognostics data simulator. A set of GPSs can be used to generate a fleet-based prognostics model.

APPENDIX A

1. Simulation of Prognostics Data

```
# Importing the libraries
import openpyxl
import numpy as np
import scipy
import scipy.stats
import random
import math
import statistics
import copy
import xlrd
import serial as ser
import time
import struct
import subprocess
import numpy as np

# Provide the path of file containing sample datasets.
path = "C:\\Users\\hp\\Desktop\\Data.xlsx"
wb = openpyxl.load_workbook(path)
# Sheet containing the data(Refer appendix C for data format)
#Sheet(0) = Data
#Sheet(1) = Parameters
#Sheet(2) =Simulation
sheet = wb.worksheets[0]
# Remove old sheets in workbook
Remove1 = wb['Parameters']
Remove2 = wb['Simulation']
```

```
wb.remove(Remove1)
wb.remove(Remove2)
wb.save(path)
# Create new sheets for storing data
wb.create_sheet('Parameters',1)
wb.create_sheet('Simulation',2)
sheet1 = wb.worksheets[1]
sheet2 = wb.worksheets[2]

#Array containing single asset datapoints.
column_vals = []
#Array containing the threshold values of each asset.
column_max = []
#Array containing the number of timestamps each sample dataset got.
life = []

# Extraction of life and threshold value from sample dataset.
for i in range(1,sheet.max_column+1):
    for j in range(1,sheet.max_row+1):
        if sheet.cell(j,i).value != None:
            column_vals.append(sheet.cell(j,i).value)

    life.append(len(column_vals)-1)
    column_max.append(column_vals[-1])
    column_vals.clear()
print(column_max)
# array containing RTF data of prognostics datasets
print(life)
```

```
#scipy.stats.weibull_min.fit provides the parameters of 2 parameter we
ibull distribution.

x = scipy.stats.weibull_min.fit(life,floc = 0)

print(x)


#write in excel sheet(1)

Titles1 = ['rand()', 'eta', 'beta', 'TTF',  'Random start mean', 'Threshold',
'k', 'Jump probability']

for i in Titles1:

    sheet1.cell(1,Titles1.index(i)+1).value = i


#Parameters of weibull distribution

sheet1.cell(2,1).value = random.uniform(0,1)

sheet1.cell(2,2).value = x[2]

sheet1.cell(2,3).value = x[0]

c = random.uniform(0,1)

power = (1/x[0])

eta = x[2]

beta = x[0]

d = (math.log(c))


#Time to failure of simulated dataset

TTF = math.ceil(x[2]  * (-d) ** (power))

print(TTF)

sheet1.cell(2,4).value = TTF


#Calculation of threshold point

column_max_mean = statistics.mean(column_max)

column_max1_stdev = statistics.pstdev(column_max)
```

```
Sim_threshold = np.random.normal(column_max_mean,column_max1
_stdev,1)

sheet1.cell(2,6).value = float(Sim_threshold)

print(Sim_threshold)


Time = []

Sr_No  = range(1,TTF+1,1)

for n in Sr_No:

    Time.append(n)

#Calculate the origin of trajectory RSM.

First_vals = []


for i in range(1,sheet.max_column+1):

    print(sheet.cell(j,i).value)

    First_vals.append(sheet.cell(2,i).value)

First_vals = (np.asarray(First_vals))

Random_start_mean = math.ceil(np.mean(First_vals))

sheet1.cell(2,5).value = Random_start_mean

print(Random_start_mean)


k = (Sim_threshold - Random_start_mean)/(TTF**2)

sheet1.cell(2,7).value = float(k)


#Write the titles in sheet(2)

Titles3 = ['Serial number','Time','Change in mean','Noise added','Rand',
'Jump magnitude','Cumulative jump','Final parameter','Seasonality mul
t','Final para with seasonality']

for i in Titles3:

    sheet2.cell(1,Titles3.index(i)+1).value = i

Jump_probability  = 0.5
```

```
Jump = []
#Generate the simulated datasets
for i in Time:
    sheet2.cell(Time.index(i)+2,1).value = i
    sheet2.cell(Time.index(i)+2,2).value = i
    CIM = Random_start_mean + (k * i **2)
    #print(CIM)
    sheet2.cell(Time.index(i)+2,3).value = float(CIM)
    #Incorporation of noise in data
    NA = np.random.normal(CIM,CIM*0.05,1)
    sheet2.cell(Time.index(i)+2,4).value = float(NA)
    rand = random.uniform(0,1)
    sheet2.cell(Time.index(i)+2,5).value = rand
    #Incorporation of jump in data
    if rand > Jump_probability:
        a = 0
        sheet2.cell(Time.index(i)+2,6).value = a

    else:
        a = random.uniform(0,5)
        sheet2.cell(Time.index(i)+2,6).value = a

for j in Time:
    if Time.index(j) == 0:
        sheet2.cell(Time.index(j)+2,7).value = sheet2.cell(Time.index(j)+
2,6).value
    else:
        sheet2.cell(Time.index(j)+2,7).value =  sheet2.cell(Time.index(j)
+2,6).value + sheet2.cell(Time.index(j)+1,7).value
```

```
    sheet2.cell(Time.index(j)+2,8).value = sheet2.cell(Time.index(j)+2,
7).value + sheet2.cell(Time.index(j)+2,4).value


# Save the data generated
wb.save(path)


#________________________________________________________


loc = ("C:\\Users\\hp\\Desktop\\Data.xlsx")


wb = xlrd.open_workbook(loc)
sheet4 = wb.sheet_by_index(4)


#Simulated RMS values of degradation parameter
rms_gen = []
rms_inc = [1]


for i in range(sheet4.nrows):
    rms_gen.append(sheet4.cell_value(i, 0))
#Calculate he percentage change in the RMS generated after each time
stamp


for i in range(len(rms_gen)-1):
    rms_inc.append(((rms_gen[i+1] - rms_gen[i])/rms_gen[i])+1)
#print(rms_inc)
# Path of data bank of motor
loc1 = ("C:\\Users\\hp\\Desktop\\Data bank.xlsx")


wb11 = xlrd.open_workbook(loc1)
sheet11 = wb11.sheet_by_index(0)
```

```
#Array containing motor databank
rms_databank = []
#Array containing Analogue write values from 0 to 165
analog_write =[]
for i in range(sheet11.nrows):
    rms_databank.append(sheet11.cell_value(i, 1))
    analog_write.append(sheet11.cell_value(i,0))
#print(rms_databank)


motor_rms = []
first_value = rms_databank[0]
for i in range(len(rms_inc)):
    motor_rms.append(first_value*rms_inc[i])
    first_value = motor_rms[i]
#Find the nearest value of datapoint in data bank respective to simulated
data to maintain the shape of degradation trajectory
def find_nearest(array, value):
    array = np.asarray(array)
    idx = (np.abs(array - value)).argmin()
    return array[idx]


motor_rms_near_value =[]
for i in range(len(motor_rms)):
    motor_rms_near_value.append(find_nearest(rms_databank,motor_r
ms[i] ))
#print(motor_rms_near_value)
#Array containing respective AWV for generation of data using the
motor vibration RMS
analog_write_near_value=[]
for i in motor_rms_near_value:
```

```
    analog_write_near_value.append(analog_write[rms_databank.index(
i)])
print(analog_write_near_value)
analog_write_near_value = list(map(int,analog_write_near_value))
print(analog_write_near_value)
print(len(analog_write_near_value))

# Muktipler calculation to amplify the RMS of motor
multiplier=(rms_gen[0])/(motor_rms_near_value[0])

#Program for arduino to rotate motor at specific AWV from array
#Send the signals to Arduino port
ser = ser.Serial('COM3',9600)
subprocess.Popen(["python", "Motor_raw_data_aquisition.py", str(len(
analog_write_near_value)),str(multiplier)], shell=True)
time.sleep()

for a1 in analog_write_near_value:
    ser.write(struct.pack('i',a1))
    time.sleep(5)

ser.close()

#----------------------------------------------------
```

2. Collection of degradation parameter trend data from motor

```
# Importing libraries
import time
from openpyxl import Workbook
import numpy as np, csv
```

```
import pandas as pd

import nidaqmx as daq, pprint

import matplotlib.pyplot as plt

from nidaqmx.stream_readers import AnalogSingleChannelReader, An
alogMultiChannelReader

import sys


pp = pprint.pprint


fig, axs = plt.subplots()

fig.canvas.manager.show()

plt.ion()


#Sampling rate for data acquisition

sampling_rate = 2500


task = daq.Task()

task.ai_channels.add_ai_voltage_chan("cDAQ9188XT-
1ADE9F6Mod3/ai1")

task.timing.cfg_samp_clk_timing(rate = sampling_rate, sample_mode
= daq.constants.AcquisitionType.CONTINUOUS)

task.in_stream.input_buf_size = (10**7)

reader = AnalogMultiChannelReader(task.in_stream)


sample_array = np.zeros([1, sampling_rate], dtype = np.float64)

task.start()

#Importing the data from previous program

no_of_speeds= int(sys.argv[1])

multiplier= float(sys.argv[2])

dur = 3
```

```
running_time = dur * no_of_speeds
num = 0

#No of timestamps in the simulated data
num_of_files = int(running_time/dur)

t_0 = time.time() - dur - 50

#Store vibration RMS of each timestamp speed of motor.
for i in range(num_of_files):
    csvobj = open("RMS_MOTOR\\speed_" + str(i + 1) + ".csv", 'w', ne
wline = '')
    csvw = csv.writer(csvobj)
    csvw.writerow(['time', 'v_y'])

    t = np.empty(shape = (0, 0), dtype = np.float64)
    v_y = np.empty(shape = (0, 0), dtype = np.float64)

    if (time.time() - t_0) < (dur):
        time.sleep((dur) - time.time() + t_0)

    t_0 = time.time()
    t_1 = 0

    for num_sec in range(dur*i, dur*(i+1)):
        t_2 = time.time()
        plt.xlim([0, num_sec + 5])

        reader.read_many_sample(data = sample_array, number_of_samp
les_per_channel = sampling_rate)
```

```
        t = np.linspace(num_sec, num_sec + 1, sampling_rate)
        v_y = sample_array[0, :]



        t_1 = t_1 + (time.time() - t_2)
        for row_num in range(0, sampling_rate):
            csvw.writerow([t[row_num].tolist(), v_y[row_num].tolist()])



        axs.plot(t, v_y, 'k-')
        axs.set_title('Y-axis vibration')



        fig.canvas.draw()
        fig.canvas.flush_events()
        plt.pause(10**(-5))

    pp("-----Time taken in acquiring:" + str(t_1))
    pp("-----Total time taken: " + str(time.time() - t_0))
    #fig.savefig("RMS MOTOR\\speed_" + str(i + 1) + ".png")
    plt.show()
    csvobj.close()

task.stop()
task.close()

wb = Workbook()
```

```
ws = wb.active
ws.title = "rms values"

for i in range(num_of_files):
    df = pd.read_csv("RMS_MOTOR\\speed_" + str(i + 1) + ".csv")
    df['c'] = df['v_y']**2
    sum = df['c'].sum(axis = 0)
    rms = (np.sqrt(sum/float(sampling_rate*dur))) * multiplier
    #plot this rms vs time after each dur
    plt.plot(running_time, rms)
    plt.xlabel('Time')
    plt.ylabel('RMS')
    plt.title("RMS")
    plt.show()
    df['rms'] = df['v_y']*0
    df.at[i + 1, 'rms'] = rms
    del df['c']
    ws.cell(column = 1, row = i+1).value = rms


# Save the data collected into excel file nemed as rms_speed
wb.save("RMS_MOTOR\\rms_speed.xlsx")
```

APPENDIX B

Arduino Programs

1 Motor vibration RMS Databank generation

Rotate the motor at each analogue write value for 5 sec starting from 0 to 255 each and capture vibration data for each analogue write value. Convert the raw data into RMS values for each 5 seconds.

```
// connect motor controller pins to Arduino digital pins
// motor one
int enA = 10;
int in1 = 9;
int in2 = 8;
int n;
void setup()
{
  // set all the motor control pins to outputs
  Serial.begin(9600);
  pinMode(enA, OUTPUT);
  pinMode(in1, OUTPUT);
  pinMode(in2, OUTPUT);


}


void loop()
//Rotate motor in increasing AWV in step of 5 seconds
{
  for (n=1 ; n<=255 ; n+=5)
  { Serial.print(n);
    analogWrite(enA,n);
    digitalWrite(in1,HIGH);
```

```
    digitalWrite(in2,LOW);
    delay (5000);
  }
// stop the motor
    analogWrite(enA,0);
    digitalWrite(in1,LOW);
    digitalWrite(in2,LOW);
    delay (50000);



}
```

2. Rotate motor for specific speeds to generate the required trend similar to the simulated data.

```
// connect motor controller pins to Arduino digital pins
// motor one
int enA = 10;
int in1 = 9;
int in2 = 8;
int n;
void setup()
{
  // set all the motor control pins to outputs
  Serial.begin(9600);
  pinMode(enA, OUTPUT);
  pinMode(in1, OUTPUT);
  pinMode(in2, OUTPUT);


}
```

```
void loop()

// Get the command from python program to rotate motor for specific
speeds.

{if (Serial.available()>0)

  {n= Serial.read();

  if (n>0)

    {

    analogWrite(enA,n);

    digitalWrite(in1,HIGH);

    digitalWrite(in2,LOW);

    delay (5000);

    }

  else

    delay(0000);

    }

    digitalWrite(in1,LOW);

    digitalWrite(in2,LOW);


}
```

APPENDIX C

1. Hardware connections of GPS setup

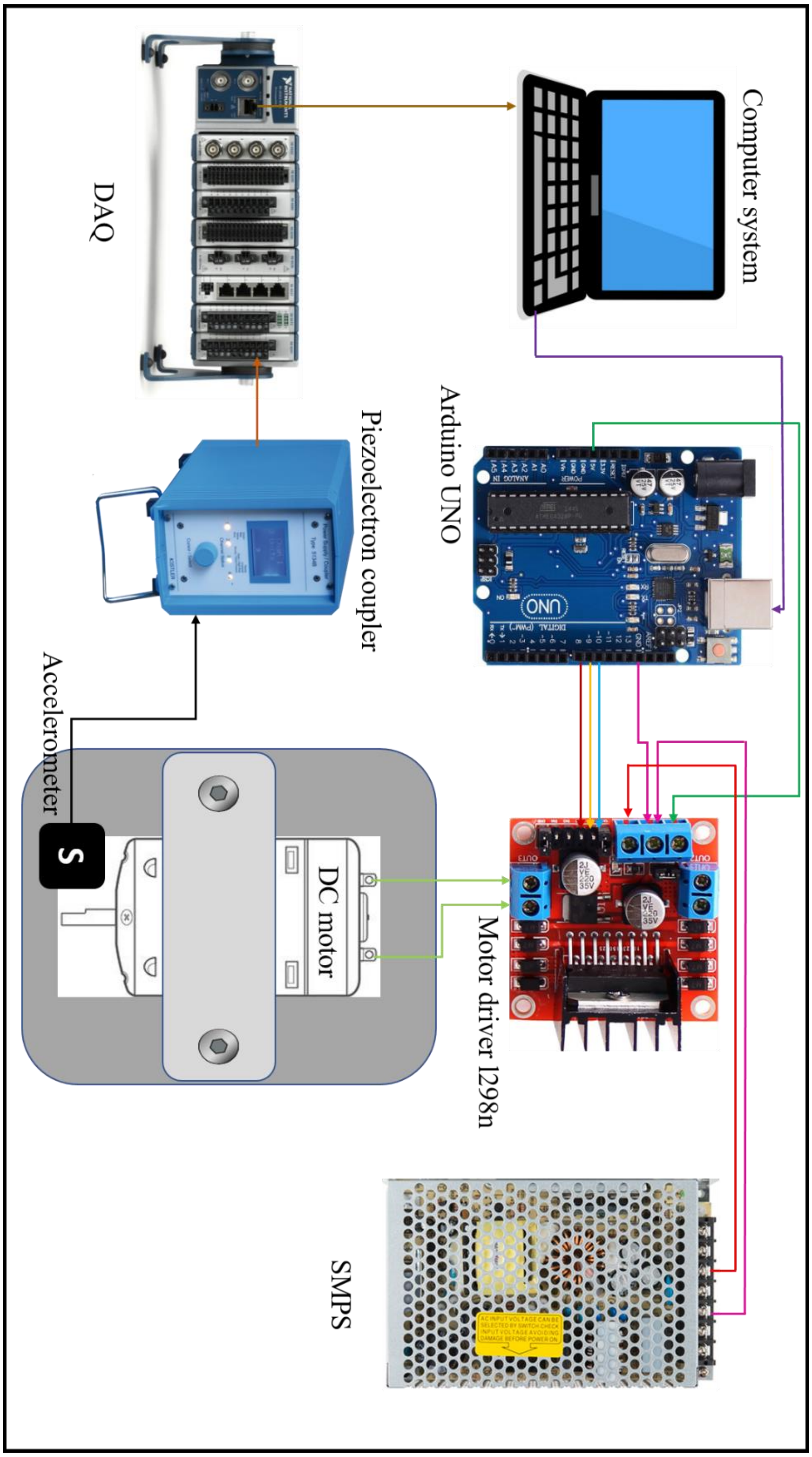

Figure 27. Hardware connection of GPS setup

Colour coding:

Table 12. Colour coding of hardware connections for GPS

Sr. No.	Color Code	Connection
1	→	Computer to Arduino
2	→	Motor driver l298n +5 power to Arduino +5v
3	→	Motor driver l298n power ground to Arduino ground, SMPS negative
4	→	Motor driver l298n +12v power to SMPS positive
5	→	Motor driver l298n Enable to Arduino 10
6	→	Motor driver l298n Input 2 to Arduino 9
7	→	Motor driver l298n Input 1 to Arduino 8
8	→	Motor driver l298n output to DC motor input
9	→	Accelerometer to coupler
10	→	Coupler to DAQ
11	→	DAQ to computer system (LAN)

Vibration RMS data of milling cutter

Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5	Dataset 6
0.009141	0.007616	0.010966	0.012058	0.011317	0.012872
0.009141	0.01114	0.01169	0.012171	0.011826	0.013328
0.009207	0.011267	0.011914	0.014541	0.012494	0.013533
0.009207	0.011309	0.013185	0.0153	0.014428	0.020338
0.00953	0.012257	0.013561	0.015944	0.025749	0.032046
0.00953	0.012392	0.014032	0.016397	0.03292	0.033739
0.011222	0.012562	0.014611	0.026946	0.03311	0.0339
0.012248	0.012708	0.018442	0.032911	0.033973	0.034986
0.012409	0.012849	0.028642	0.033481	0.037579	0.035176
0.012619	0.012966	0.03794	0.034165	0.038664	0.036783
0.012683	0.01326	0.038887	0.037765	0.042887	0.038758
0.012982	0.013664	0.05182	0.039015	0.043949	0.042045
0.013245	0.013664	0.054809	0.039332	0.126267	0.129578
0.014364	0.013815	0.117195	0.124196	0.129263	0.146192
0.01473	0.013939	0.117195	0.125688	0.139271	
0.014924	0.013939	0.117901	0.135469		
0.014931	0.013987	0.142411	0.138394		
0.014949	0.01457	0.142474	0.158433		
0.023062	0.014895	0.143952			
0.028111	0.015252	0.148532			
0.039078	0.018666	0.201612			
0.067038	0.03278				
0.078939	0.035019				
0.116845	0.036246				
0.125957	0.038736				
0.137352	0.131496				
0.146226	0.161552				
0.154995	0.163142				
0.157343	0.196304				
0.403206					

Wear data of milling cutter

Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5	Dataset 6
0.000	0.000	0.000	0.000	0.000	0.000
0.011	0.042	0.015	0.029	0.041	0.040
0.031	0.070	0.025	0.126	0.074	0.117
0.048	0.081	0.070	0.159	0.149	0.181
0.053	0.091	0.114	0.189	0.167	0.199
0.063	0.098	0.135	0.223	0.278	0.241
0.070	0.112	0.172	0.245	0.312	0.322
0.076	0.129	0.190	0.262	0.360	0.376
0.078	0.140	0.208	0.310	0.414	0.430
0.083	0.165	0.270	0.364	0.467	0.484
0.098	0.190	0.297	0.431	0.521	0.538
0.152	0.197	0.320	0.465	0.612	0.645
0.192	0.245	0.378	0.538	0.628	0.717
0.203	0.274	0.406	0.577	0.681	0.753
0.242	0.331	0.487	0.618	0.751	
0.286	0.393	0.520	0.647		
0.305	0.413	0.550	0.687		
0.329	0.436	0.565	0.763		
0.350	0.458	0.601			
0.410	0.483	0.636			
0.441	0.504	0.672			
0.467	0.529				
0.498	0.555				
0.518	0.581				
0.538	0.607				
0.579	0.633				
0.609	0.659				
0.643	0.685				
0.666	0.710				
0.685	0.791				
0.794					

REFERENCES

[1] J. B. Coble, "Merging Data Sources to Predict Remaining Useful Life – An Automated Method to Identify Prognostic Parameters," University of Tennessee, Knoxville, 2010.

[2] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao and D. Siegel, "Prognostics and health management design for rotary machinery systems—Reviews,methodology and applications," *Mechanical Systems and Signal Processing,* vol. 42, pp. 314-334, 2014.

[3] K. Goebel, A. Saxena, M. Daigle, J. Celaya and I. Roychoudhury, "Introduction to Prognostics," in *First European Conference of the Prognostics and Health Management Society*, Dresden, Germany, 2012.

[4] H. M. Elattar, H. K. Elminir and A. M. Riad, "Towards online data-driven prognostics system," *Complex & Intelligent Systems,* vol. 4, p. 271–282, 2018.

[5] N. Chen and K. L. Tsui, "Condition monitoring and remaining useful life prediction using degradation signals: revisited," *IIE transaction ,* vol. 45, no. 9, 2013.

[6] Y. Li, S. Jianming, W. Gong and L. Xiaodong, "A data-driven prognostics approach for RUL based on principle component and instance learning," in *International conference on Prognostics and health management (ICPHM),* Beijing, 2016.

[7] J. Kamran, G. Rafael and Z. Noureddine, "State of the art and taxonomy of prognostics approaches, trends," *Mechanical Systems and Signal Processing ,* vol. 94, no. 15, pp. 214-236, 2017.

[8] J. P. MacDuffie and T. Fujimoto, “Why dinosaurs will keep ruling the auto industry,” *Harvard Business Review,* pp. 23-25, 2010.

[9] K. L. Tsui, N. Chen, Q. Zhou, Y. Hai and W. Wang, “Prognostics and Health Management: A Review on Data Driven Approaches,” *Hindawi Publishing Corporation Mathematical Problems in Engineering, Article ID 793161,* 2015.

[10] K. Holmberg, K. Komonen, P. Oedewald, M. Peltonen, T. Reiman, V. Rouhiainen, J. Tervo and P. Heino, “Safety and reliability technology review,” Res Rep BTUO43-031209. VTT Industrial, 2004.

[11] T. Dong, R. T. Haftka and H. K. Nam, “Advantages of Condition-Based Maintenance over Scheduled Maintenance using Structural Health Monitoring System,” in *System Reliability,* Intech Open, 2019.

[12] Y. Peng, M. Dong and M. J. Zuo, “Current status of machine prognostics in condition-based maintenance: a review,” *Int J Adv Manuf Technol,* vol. 50, p. 297–313, 2010.

[13] A. Gillespie, “Condition Based Maintenance: Theory, Methodology, & Application,” in *Reliability and Maintainability Symposium,* Tarpon Springs, FL, 2015.

[14] D. O. Gray and D. Rivers, “Measuring the Economic Impacts of the NSF Industry/University Cooperative Research Centers Program: A Feasibility Study,” Industry-University Cooperative Research Center Program, National Science Foundation, Arlington, Virginia, 2012.

[15] M. Wiseman, “A History of CBM (Condition Based Maintenance),” 2006. [Online]. Available: http://www.omdec.com/moxie/Technical/Reliability/a-history-of-cbm.shtml. [Accessed 2006].

[16] S. Ferreiro, E. Konde, S. Fernández and A. Prado, “INDUSTRY 4.0: Predictive Intelligent Maintenance for Production Equipment,” in *EUROPEAN CONFERENCE OF THE PROGNOSTICS AND HEALTH MANAGEMENT SOCIETY*, Bilbao, Spain, 2017.

[17] V. M. Catterson, J. J. A. Costello, W. G. M. S. D. J. McArthur and C. J. Wallace, “Increasing the Adoption of Prognostic Systems for Health Management in the Power Industry,” *Chemical Engineering Transactions,* vol. 33, pp. 271-276, 2013.

[18] F. Elasha, S. Shanbr, X. Li and D. Mba, “Prognosis of a Wind Turbine Gearbox Bearing Using Supervised Machine Learning,” *Sensors,* vol. 19, no. 14, 2019.

[19] S. Mathew and M. Pecht, “Prognostics of Systems: Approaches and Applications,” in *Condition monitoring and diagnostic engineering management.*, Maryland, 2011.

[20] K. K. Reddy, “Applying Deep Learning for Prognostic Health Monitoring of Aerospace and Building Systems,” in *Knowledge discovery and data mining*, San francisco, 2016.

[21] D. Luchinsky, S. Osipov, V. Smelyanskiy, C. Kiris, D. Timucin and S. H. Lee, “In-Flight Failure Decision and Prognostics for the Solid Rocket Buster,” *AIAA 43rd AIAA/ASME/SAE/ ASEE Joint Propulsion Conference and Exhibit, Cincinnati, OH, July 8-11,* 2007.

[22] G. D. Ranasinghe and A. K. Parlikad, “Generating Real-valued Failure Data for Prognostics Under the Conditions of Limited Data Availability,” in *International Conference on Prognostics and Health Management (ICPHM),* 2019.

[23] A. Dawn, C. Joo-Ho and H. K. Nam, “Prediction of remaining useful life under different conditions using accelerated life testing

data," *Journal of Mechanical Science and Technology*, vol. 32, no. 6, pp. 2497-2507, 2018.

[24] H. Skima, K. Medjaher and N. Zerhouni, "Accelerated life tests for prognostic and health management of MEMS devices," in *EUROPEAN CONFERENCE OF THE PROGNOSTICS AND HEALTH MANAGEMENT SOCIETY 2014*, Nantes, France, 2014.

[25] A. K. Jardine, D. Lin and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing* , vol. 20, no. 7, p. 1483–1510, 2005.

[26] T. Wang, J. Yu, D. Siegel and J. Lee, "A Similarity-Based Prognostics Approach for Remaining Useful Life Estimation of Engineered Systems," in *INTERNATIONAL CONFERENCE ON PROGNOSTICS AND HEALTH MANAGEMENT*, Denver, Colorado, USA, 2008.

[27] A. Dawn, H. K. Nam and C. Joo-Ho, "Options for Prognostics Methods: A review of data-driven and physics-based prognostics," in *Annual Conference of the Prognostics and Health Management Society*, New Orleans, 2013.

[28] J. Igba and K. Alemzadeh, "Analysing RMS and peak values of vibration signals for condition monitoring of wind turbine gearboxes," *Renewable Energy* , vol. 91, pp. 90-106, 2016.

[29] C. Wahyu, "Vibration and acoustic emission-based condition monitoring and prognostic methods for very low speed slew bearing," School of Mechanical, Materials and Mechatronic Engineering, University of Wollongong, Wollongong, 2015.

[30] E. Bechhoefer and M. Revor, "Rotor track and balance cost benefit analysis and impact on operational availability," American

Helicopter Society 60th Annual Forum, Baltimore, MD, 2004, Baltimore, 2018.

[31] P. Chauhan and Q. Yu, "Use of Temperature as a Health Monitoring Tool for Solder Interconnect Degradation in Electronics," in *Prognostics & System Health Management Conference PHM-2012*, Beijing, 2017.

[32] A. R. Kumar and V. Krishnan, "A Study on System Reliability in Weibull Distribution," *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering*, vol. 5, no. 3, pp. 38-41, 2017.

[33] L. P. F. Garcia, d. Carvalho Andre and A. C. Lorena, "Noisy Data Set Identification," in *International Conference on Hybrid Artificial Intelligence Systems*, Asturias, 2018.

[34] A. M. DAVEY and B. E. FLORES, "Identification of Seasonality in Time Series: A Note," *Mathl. Comput. Modelling* , vol. 18, no. 6, pp. 73-81, 1993.

[35] K. Wallis, "Seasonal adjustment and relations between variables," *J. American Stat. Assoc*, vol. 69, no. 34, pp. 18-31, 1974.

[36] M. R. Khan, K. Mohd and P. Kumar, "Speed Control of DC Motor by using PWM," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 5, no. 4, pp. 307-309, 2016.

[37] J. Lee, C. Jin, Z. Liu and H. D. Ardakani, "Introduction to Data-Driven Methodologies for Prognostics and Health Managemen," *Probabilistic Prognostics and Health Management of Energy Systems*, pp. 10-33, 2018.

[38] W. Caesarendra and T. Tjahjowidodo, "A Review of Feature Extraction Methods in Vibration-Based Condition Monitoring and

Its Application for Degradation Trend Estimation of Low-Speed Slew Bearing," *Machines*, vol. 5, no. 21, pp. 1-28, 2017.

[39] A. Huang, "Similarity Measures for Text Document Clustering," in *New Zealand Computer Science Research Student Conference*, Christchurch, 2008.