IDENTIFICATION OF FLOW BLOCKAGE LEVELS IN CENTRIFUGAL PUMP BY MACHINE LEARNING

By:

NG WOON LI

(Matric No: 137854)

Supervisor:

Dr. Muhammad Razi Bin Abdul Rahman

July 2021

This dissertation is submitted to
Universiti Sains Malaysia
As partial fulfillment of the requirement to graduate with honors degree in
BACHELOR OF ENGINEERING (MECHANICAL ENGINEERING)

School of Mechanical Engineering
Engineering Campus
Universiti Sains Malaysia

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed: ……… ……………………… (NG WOON LI)

Date: ……………12/7/2021……………………………………

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by giving explicit references. Bibliography/references are appended.

Signed: ……… ……………………… (NG WOON LI)

Date: ……………12/7/2021……………………………………

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for interlibrary loan, and for the title and summary to be made available outside organizations.

Signed: ………… …………………… (NG WOON LI)

Date: ……………12/7/2021……………………………………

ACKNOWLEDGEMENT

First and foremost, I would like to express my sincere gratitude to the supervisor of my final year project, Dr. Muhammad Razi bin Abdul Rahman for guiding me throughout the project. It is my honour to have his timely advice, dedication, supports and assistance in this project that encourage me to do better. I would like to acknowledge the hard work by Dr. Muhammad Fauzinizam Bin Razali as the course coordinator that scheduled seminars and plan the course structure.

Besides, I would like to thank School of Mechanical Engineering for providing me the precious opportunities to gain knowledge and hands-on experience throughout my four years of studies. I am convinced that all the technical skills and enlightenments are applicable in future career. Other than that, I am extremely thankful for the experimental rig available in Vibration Lab which allowed me to complete the studies of my project. In addition, I thank profusely the research officer in Vibration Lab, Mr Wan Mohd Amri Wan Mamat Ali who assist and guide me through the experiment.

Finally, the guidance and advice gave by the lecturers and examiner is well appreciated as it gave me confidence to conduct my project better. My thanks and appreciations also go to the course mates who giving me mentally supports during the project execution.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

ML	Machine learning
TDA	Time domain analysis
FDA	Frequency domain analysis
SVM	Support Vector Machine
FFT	Fast Fourier Transform
GUI	Graphical User Interface
RMS	Root Mean Square
KNN	K-Nearest Neighbour
MATLAB	Matrix Laboratory
AI	Artificial Intelligence
MDP	Markov Decision Process
NPSHa	Net Positive Suction Head available
NPSHr	Net Positive Suction Head required
MFS	Machine Fault Simulator
DAQ	Data Acquisition System
EDT	Electric Diagnostic Technique
LabVIEW	Laboratory Virtual Instrument Engineering Workbench
CFD	Computational Fluid Dynamics
EDA	Exploratory Data Analysis
RBF	Radial Basis Function
ANN	Artificial Neural Network
PCA	Principal Component Analysis
GA	Genetic Algorithm
ROC	Receiver Operating Characteristics

LIST OF APPENDICES

KLASIFIKASI UNTUK TAHAP PENYUMBATAN SALURAN PAM SENTRIFUGAL DENGAN PEMBELAJARAN MESIN

ABSTRAK

Projek ini fokus dalam pemantauan penyumbatan saluran pam sentrifugal dengan analisis getaran dan akustik. Penyumbatan dalam saluran masuk pam akan menyebabkan peronggaan atau kerosakan bahagian mekanikal yang akan meningkatkan kos penyelenggaraan. Pembelajaran mesin boleh digunakan sebagai langkah pencegahan untuk mengesankan penyumbatan saluran masuk pam pada tahap permulaan. Tujuan penyelidikan ini adalah untuk menghasilkan model pembelajaran mesin untuk klasifikasi tahap penyumbatan aliran dalam pam sentrifugal dengan menggunakan ciri-ciri signifikan dari analisis getaran dan akustik. Data mentah telah dikumpulkan dengan menggunakan sistem data pemerolehan Siemens LMS SCADAS Mobile. Selanjutnya, ciri statistik diekstraksi dan dilatih oleh beberapa model klasifikasi yang terdapat dalam Classification Learner, MATLAB R2020b dan disahkan dengan ‘5-fold cross-validation’. RMS, varians dan kurtosis daripada data akustic dalam domain masa didapati mempunyai potensi ramalan yang tinggi berdasarkan keputusan ujian Chi-Square. Selain itu, Ensemble Bagged Tree and Support Vector Machine (SVM) dengan kernel kubik telah mencapai ketepatan yang tertinggi, adalah 95.8% dan 94.4%. Model SVM dengan kernel kubik diutamakan kerana masa latihan yang diambil adalah lebih rendah daripada Ensemble Bagged Tree yang lebih kompleks. Oleh itu, model SVM dengan kernel kubik dieksport sebagai fungsi MATLAB untuk membuat ramalan untuk data baru. Selain itu, hasil model Linear SVM telah disahkan dengan menggunakan pakej klasifikasi di Google Colaboratory. Beberapa penambahbaikan yang berpotensi telah disenaraikan dalam laporan ini untuk meningkatkan ketahanan dan ketepatan model pembelajaran mesin.

IDENTIFICATION OF FLOW BLOCKAGE LEVELS IN CENTRIFUGAL PUMP BY MACHINE LEARNING

ABSTRACT

The current project focused on the flow blockage monitoring of centrifugal pump by vibration and acoustic analysis. The blockage of the pump inlet could result in cavitation or mechanical parts breakdown which would increase the maintenance cost. Machine learning can be used as a preventive measure to detect the blockage in the pump inlet at inception level. The purpose of this research is to develop an effective machine learning model for the classification of flow blockage levels in the centrifugal pump by using the statistically significant features from vibration and acoustic analysis. Raw data are collected by using Siemens LMS SCADAS Mobile data acquisition system. Furthermore, the statistical features are extracted and trained by multiple classifiers available in Classification Learner, MATLAB R2020b and validated by 5-fold cross-validation. RMS, variance and kurtosis of acoustic signals in time domain have strong predictive potential based on the Chi-Square test which ranked the predictive power of features for classification. Besides, Ensemble Bagged Tree and Support Vector Machine (SVM) with cubic kernel has achieved the highest accuracy, which are 95.8% and 94.4%, respectively. SVM model with cubic kernel is preferable as the training time taken is relatively lower than Ensemble Bagged Tree due to the ensemble algorithms are more complex. Hence, the SVM model with cubic kernel is exported as a MATLAB function to make predictions for the new data. Besides, the result of the Linear SVM model is validated by using the classification packages in Google Colaboratory. Several potential improvements are recommended in this report to increase the robustness and accuracy of the machine learning model.

CHAPTER 1

INTRODUCTION

1.1 Project Background

Centrifugal pump is a mechanical device commonly used to convert rotational kinetic energy to hydrodynamic energy to transfer fluid. It is popular in the industry as it is known for high efficiency with low power consumption. However, the problems arise within the machine will result in a drop in the flow rate of the pipeline. If the flow blockages of the pump are not timely treated, pump failure might occur, and the centrifugal pump must be replaced.

Therefore, experts generally agree that monitoring the operation of the centrifugal pump is significant to diagnose the faulty mode prior to the pump failing catastrophically to reduce downtime (Tiwari, 2017). The vibration and acoustic signals contain the characteristics of failure present in rotating machines. Vibration and acoustic monitoring have been proven to be effective in detecting faults within the centrifugal pump. Pump manufacturers start to provide onboard sensors on the pump to ease the data analysis process as some of the pumps are in an inaccessible environment such as water supply or sewage industries (Cao & Yuan, 2020). Accelerometers and microphone are used to collect the vibration and acoustic data respectively in this project.

Therefore, preventive measures shall be taken to minimize the possibilities of sudden breakdown and to reduce maintenance costs as frequent maintenance might not be required. Machine learning can be implemented to study the characteristics of vibration and acoustic signals to identify the blockage condition of the centrifugal pump inlet.

Machine learning approaches consist of three major categories which are supervised learning, unsupervised learning and reinforcement learning. In this project, supervised learning is utilized where the features extracted from vibration and acoustic datasets with labelled blockage conditions are fed into the machine learning model. Classification of the flow blockage conditions of centrifugal pump can be done with the classifier such as Support Vector Machine (SVM), K-Nearest Neighbor (KNN), etc. The expected output of the classifier is to classify the dataset into several levels of flow blockage. The classifiers are evaluated by K-fold cross-validation where the dataset are split into K subsets, one of which will be used for testing, and other subsets are training sets. The validation process is repeated for K rounds with each subset taking turns to be testing set.

1.2 Problem Statement

Blockage in the inlet of a centrifugal pump is a significant early sign of other issues that can cause cavitation in centrifugal pump. Clogging of the pump inlet might result in a smaller area of fluid passage and higher fluid velocity. It will cause the fluid pressure to become lower, and cavitation bubbles will form if the pressure of fluid falls below the saturated vapour pressure. Besides, it generates loud noise and vibration due to the unstable flow of fluid.

Based on the research works discussed by Isermann, unplanned repairs of centrifugal pump due to unexpected failure occurred within a mean of nine months. 59% of the pumps in the chemical industry or water treatment plant operate continuously, and maintenance are done every three months (Isermann, 2011). Furthermore, centrifugal pump inspection procedures are highly dependent on the instrument, such as the flow transducer. However, the outlet pressure measurement only will be alerted when large deviations occur, and conventional instruments do not diagnose the source of the problem, such as flow blockage, bearing failures etc. The current general supervision method does not allow early detection of minor faults and unable to diagnose the cause of the problems.

1.3 Objectives

This project aims to achieve the following objectives:

1. To determine the statistically significant features of flow blockages in centrifugal pump based on vibration and acoustic signals.
2. To develop an effective machine learning model for the classification of flow blockage levels with the extracted statistical features by comparing the results of various classifiers available in MATLAB Classification Learner.
3. To validate the results from MATLAB Classification Learner with the classification packages in the Google Colaboratory machine learning platform.

1.4 Scope of Project

Flow blockages of the centrifugal pump were investigated by using the experimental rig available in the VibrationLab laboratory. Two accelerometers, a microphone and a tachometer are deployed to collect vibration, acoustic and rotational speed of the centrifugal pump. The accelerometers were allocated in the vertical axis and radial axis to discover the direction of vibration induced by pump blockage. The acoustic and vibration data were collected by using Siemens LMS SCADAS data acquisition system and the LMS Test Xpress software. Six blockage conditions of a centrifugal pump were discussed in this project which is ‘no blockage’, ‘inception’, ‘development’, ‘mild’, ‘severe’ and ‘mild and unstable’. The blockage of the pump inlet is simulated by altering the inlet opening valve. Besides, the experiment for each condition was repeated for four times, yielding a total of 24 sets of data. Then, the obtained data are then downsized three times with a factor of 5, 10 and 15. Therefore, there is a total of 72 sets of data available for the supervised machine learning process. The datasets are transformed into frequency domain by using FFT. There is a total of 30 statistical features extracted from the time domain and frequency domain of raw data for the three sensors. Chi-Squared test is used to rank the features according to their predictive power. Feature selection is used to reduce the models’ complexity and training time. Data visualization, features extraction and data training with various classifiers are carried out by using MATLAB R2020b. Classification Learner, one of the GUI in MATLAB R2020b is used to execute the machine learning process in this project.

The limitations of the project are the data collection is only limited to one centrifugal pump. Hence, the statistical features studied are only based on a single type of centrifugal pump. Besides, the project only concentrates on blockage of the inlet of centrifugal pump. Therefore, other faults of the pump that might be contributed to the noise and vibration are neglected.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction to Centrifugal Pump

Centrifugal pump as shown in Figure 2.1, are ubiquitous in the industry as it is used in building ventilation system, cooling tower and others industrial applications. Fluid enters axially through the hollow middle portion of the pump known as the impeller eye after it encounters the rotating blades and exits axially. It acquires tangential and radial velocity by momentum transfer with the impeller blades and acquires additional radial velocity which is centrifugal forces. The fluid will leave the impeller by gaining both speed and pressure as it flung radially to the volute. The volute is designed to be a snail-shaped diffuser in order to decelerate the fast-moving fluid leaving the trailing edges of the impeller blade and increase the fluid pressure to direct the flow to the discharge pipe (Cengel & Cimbala, 2013).

There are three types of centrifugal pumps which is forward-inclined blades, backward-inclined blades, and radial blades. However, the backward-inclined blades have the highest efficiency as it requires the least amount of turning the fluid flows in and out.

Centrifugal pump can be used to move fluid around the system or transfer fluid from one tank to another. Fluids such as water, sewage, and petroleum are transferred using the centrifugal pump in the industry.

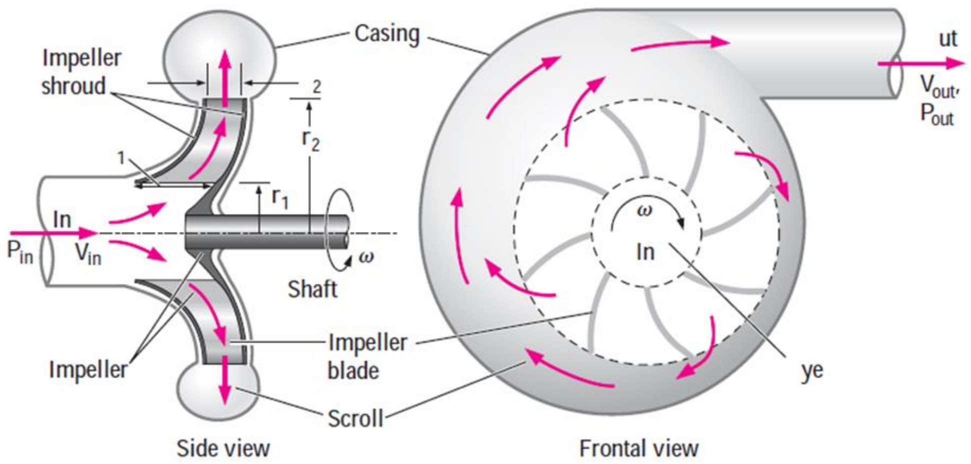

Figure 2.1: Centrifugal Pump (Cengel & Cimbala, 2013).

2.1.1 Flow Blockage in Centrifugal Pump

In industrial applications, flow blockage in a centrifugal pump occurs due to the deposit of organic material or through chemical reactions at the pump inlet or outlet, resulting in less pump efficiency. It might cause a sudden breakdown of the pump due to the high temperature generated by the pump's motor (Isermann, 2011). The pump fluid might contain materials such as rags that clog the passage which might decrease the flow rate. Flow blockage of the suction pipe will cause restriction flow of fluid and eventually lead to cavitation in the centrifugal pump.

There are several reasons that causes flow blockage in a centrifugal pump as it also depends on the location and pumping fluid. Objects like loose rags, natural detritus, grass, etc. can contribute to the blockage of a pump. Other than that, pipe liner disintegration and air trapped at the suction inlet of the pump will lower the flow capacity as well (Panda, Rapur, & Tiwari, 2018). Furthermore, flow blockage will result in impending or development of cavitation on the impeller vane. The vibration of the axial shaft, impeller shroud cracking, excessive noise and vibration are the additional possible consequences (Fraser, 1982).

2.1.2 Cavitation of Centrifugal Pump

Cavitation of centrifugal pump occurs when the static pressure is lower than the vapour pressure of the liquid. The air bubbles will expand as it approaches boiling point and collapses abruptly and leads to damage to the metal part, especially the impeller blades resulting in crackling noise and vibration. It usually occurrs when the flow rate is higher than the designed flow rate, reduction of suction pressure or rising of suction temperature (McKee et al., 2011). Cavitation is one of the challenging fluid flow abnormalities that leads to detrimental effects. Cavitation causes degradation of pump performance and leads to noise and vibration. There are several types of cavitation such as shear cavitation, attached cavitation, travelling cavitation and vortex cavitation. Cavitation occurs in several situations such as the wall geometry that gradually decreases in flow area will impose high velocity and pressure drop, high unsteadiness characteristics of a particular flow which yield pressure drop in fluid flow, etc. (Maxime Binama, Alex Muhirwa, 2016)

Net positive suction head (NPSH) curve is designed by the pump manufacturer where cavitation usually happened during the actual NPSH (NPSHa) is lower than the required NPSH (NPSHr). The actual NPSH is calculated by the characteristics found in the pump's suction nozzle, whereas the required NPSH is the net positive suction head required to avoid cavitation. According to Figure 2.2, the centrifugal pump shall operate within the flow rate before the interception of NPSHA and NPSHR to prevent cavitation.

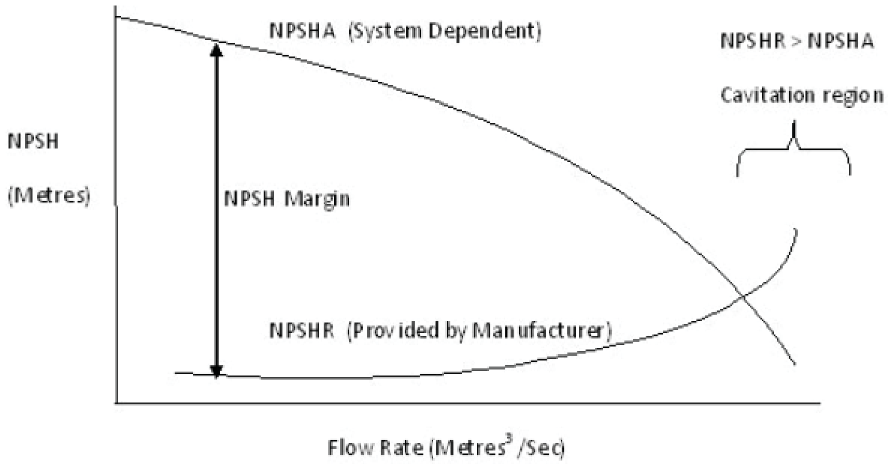

Figure 2.2: Graph of NPSH against flow rate. (McKee et al., 2011)

There are a few symptoms that indicate cavitation is taking place within the centrifugal pump. The bubbles collapsing at high-pressure areas exert enormous local stresses on the pump surface that results in pitting. Besides, the cavities collapsing under high pressure causes sharp crackling sound and vibration. Furthermore, the vapour bubbles in the passages around the impeller impede the flow of fluid. As a result, the discharge flow rate is reduced, causing low pump efficiency. The author also mentioned some corrective methods that could be made to prevent cavitation damage, such as replacing the impeller with cavitation resistant material like stainless steel. Besides, redesign the impeller's geometry to reduce losses and improve flow characteristics or place the inducer at the suction part of the pump to increase the pressure before fluid reaches the impeller (McKee et al., 2011).

Cavitation issue is not investigated in this study as this project focus on the flow blockage in a centrifugal pump, which is an early sign of cavitation. However, the concepts of cavitation are deployed to explain the observations and discoveries of the results.

2.2 Experimentations with Centrifugal Pumps

Based on Panda (2018) research, the centrifugal pump is mounted on the fixed base of Machine Fault Simulator (MFS) with an operating temperature of 299K. Water is used as the pumping fluid, and the experimental setup is designed in a closed loop. There are two tri-axial accelerometers attached to the pump casing and bearing housing. The vibration data are collected at 20kHz sampling frequency. Five different blockage conditions are simulated by manually modulating the pump inlet valve, as depicted in Figure 2.3. Based on the observation of this research, cavitation occurred when the blockage is 50% and 66.6% as the air bubbles started to appear. The cavitation effect rises as the blockage level increases. Besides, the pressure condition of the pump discharge outlet recorded shows the pipe pressure decreased as the cavitation level increases. Similarly, another researcher also creates pipe blockage by adjusting the inlet valve (Irfan, Alwadie, & Glowacz, 2019).

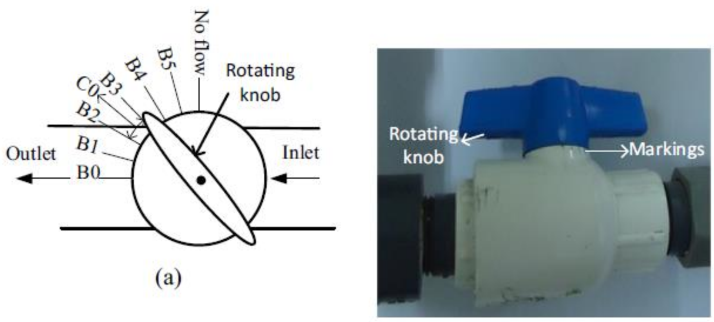

Figure 2.3: Manual modulating valve of the pump inlet. (Panda et al., 2018)

2.2.1 Position of Accelerometer

Accelerometers are widely used in numerous experiments as vibration sensors. It can operate in a wide range of frequencies, reliable for vibration measurement and easy to install on the machine. The optimal mounting position of the accelerometer on the centrifugal pump is axial to the suction position. The vibration signals will be preprocessed by using Data Acquisition System (DAQ) to amplify and filter the signal with a suitable sampling rate. It is found that the RMS and peak values being the best parameters of time domain for detecting cavitation (AL Tobi & Al Sabari, 2016). Besides, the tri-axial accelerometer can be installed on the pump casing and bearing housing to capture vibration signals. (Panda et al., 2018).

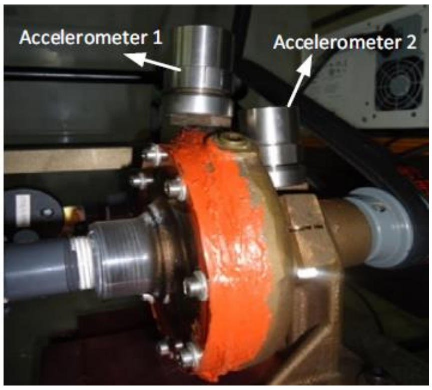

Figure 2.4: Position of accelerometers on centrifugal pump. (Panda et al., 2018)

2.3 Diagnosis of Faults in Centrifugal Pump

Condition monitoring (CM) is usually used to predict cavitation with the parameters of the pumping process, especially the suction pressure. However, some implementations of CM using pressure sensors are intrusive since the sensors are in contact with the pumping fluid. Besides, tapping the sensors in the pipe might lead to leakage issues. Therefore, pressure sensor is disqualified, especially for dangerous fluid. Alternatively, vibration sensor has significant advantages as a non-intrusive sensor, and it can contribute information for a wide range of rotating machinery faults (Kléma, Flek, Kout, & Nováková, 2005).

Rotating machinery will generate specific oscillation from various components such as the rotating shafts, blades or rotor. The oscillation will be more pronounced when there are turbulent flows of fluid or cavitation. Besides, the arise vibration might depend on the flow rate and rotational speed as well. Therefore, defects in the centrifugal pumps can be detected by using sensors such as oscillation velocity sensors, oscillation accelerometers or piezoelectrical sensors (Isermann, 2011).

Aside from that, electric diagnostic technique (EDT) does not require additional sensors as it uses the sensors that already installed in the centrifugal pump to monitor the motor line current and voltage. Laboratory Virtual Instrument Engineering Workbench (LabVIEW) software can be used to measure the three-phase line current and transform it to two-phase d-q current (LabVIEW TM Getting Started with LabVIEW Getting Started with LabVIEW, 2013). The statistical characteristics of the d-q current plot can be employed to categorize the pump faults. The flow chart of the EDT condition monitoring system is explained in Figure 2.5. The d-q patterns are unique for a healthy pump, impeller faults, and pipe blockages. Thus, it can be a visual indicator for the pump health's condition (Irfan et al., 2019). Computational Fluid Dynamics (CFD) has been employed to predict the performance of centrifugal pump with applied mathematics, physics, and software to visualise the flow within the pump and how it affects the machinery within the pump. This method yields high accuracy, time and resources saving and provides flow visual ability (Maxime Binama, Alex Muhirwa, 2016).

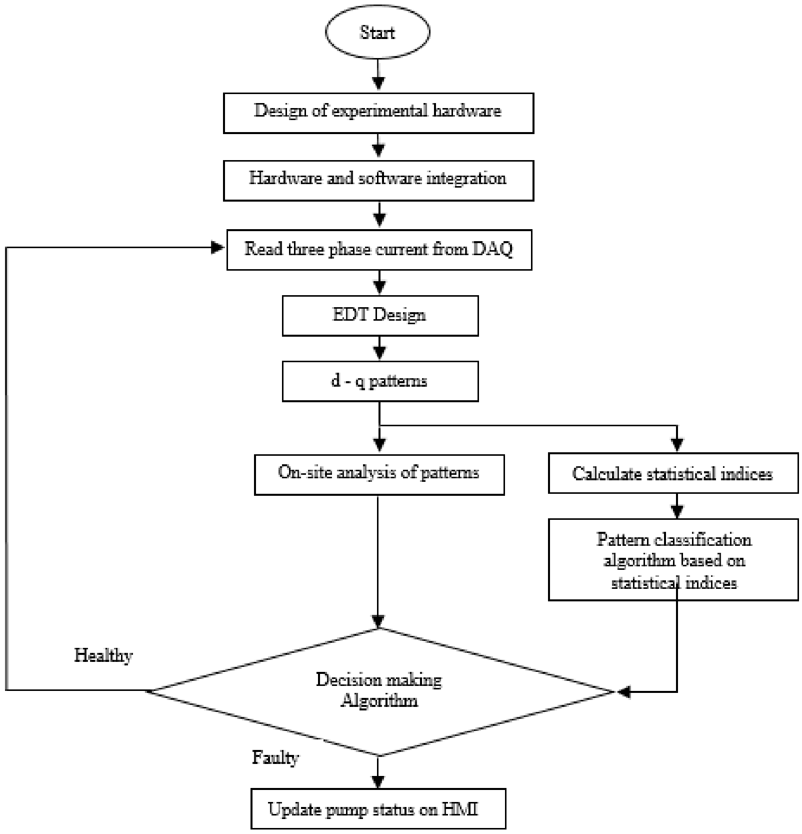

Figure 2.5: Flow chart of EDT condition monitoring system. (Irfan et al., 2019)

2.3.1 Diagnosis Method by Machine Learning

Machine learning (ML) is an Artificial Intelligence (AI) application that uses data analytical tools to learn and identify features from data. ML can also be defined as a computational method using experience to improve performance or to make accurate predictions (Mohri, Rostamizadeh, & Talwalkar, 2018). It can make a decision with minimal human intervention. Machine learning consists of three major categories which are supervised learning, unsupervised learning and reinforcement learning. Supervised learning uses a dataset comprising both inputs and desired output to train the machine learning model to predict the outcome of new input. Classification and regression are categorised as the supervised learning algorithm. Unsupervised learning uses data with only input and identifies the pattern of the data set by clustering. Besides, reinforcement learning employs a rewards system where it will be rewarded by interacting with its environment. For instance, the Markov decision process (MDP) is commonly used in autonomous vehicles or playing game against human opponent (Guan, Li, Duan, Wang, & Cheng, 2018).

According to German Fachgemeinschaft Pumpen, VDMA, most of the water treatment plant in the chemical industry has 59% of its centrifugal pumps operating continuously, whereas 19% running daily and 22% operates for a short period. The pumps are inspected every three months in average. However, unplanned repairs are needed within a mean of nine months. The conventional supervision methods are done by monitoring the discharge outlet pressure and supervise the inlet pressure. The operator will only be notified when a significant failure occurred. When the discharge outlet pressure measured at constant rotational speed has a significant deviation compared to the normal operating pressure or inlet pressure that exceeds the Net Positive Suction Head (NPSH), it indicates that there are failures occurred within the centrifugal pump. These traditional supervision approaches do not allow early detection of minor faults and show the diagnosis of the source of failure. The pump has to be replaced by the time the failures are found (Isermann, 2011).

Hence, machine learning algorithms are integrated into the condition monitoring of centrifugal pumps to reduce the maintenance cost and avoid sudden breakdown or sudden repair. Machine learning algorithms is a computational method that can quickly detect faults in the centrifugal pump. The types of machine learning algorithms are shown in Figure 2.6. (Dutta et al., 2018).

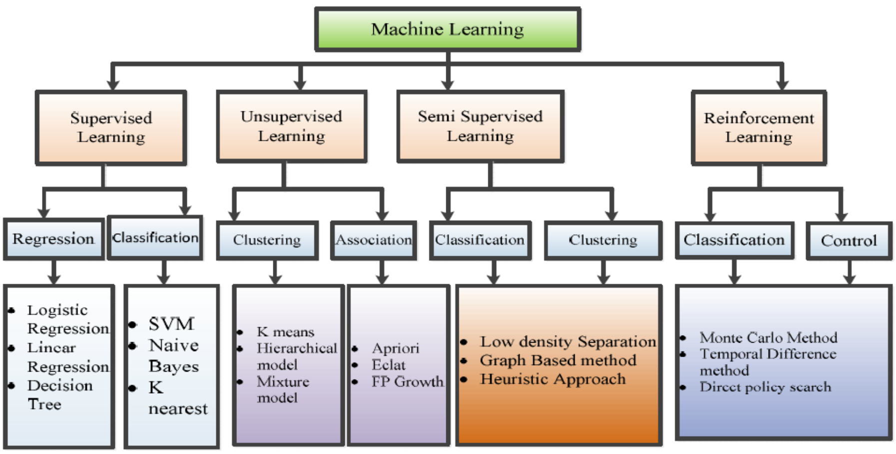

Figure 2.6: Machine learning algorithms. (Dutta et al., 2018)

Today, there are numerous research works regarding constant monitoring of the centrifugal pump by using vibration sensors or pressure gauges in the pumps. Based on Liang Dong (2019) research, wavelet packet decomposition (WPD), principal component analysis (PCA), and radial basic function (RBF) neural network are deployed in cavitation identification of centrifugal pump. The energy coefficient of each node of the time-frequency domain analysis of interior flow-borne noise is to identify the best decomposition frequency band. RBF neural network is used as the algorithm to identify three cavitation statuses: non-cavitation, inception cavitation and serious cavitation. The model is able to achieve an accuracy of 98.2% with minor classification error. PCA is deployed to improve the recognition speed by reducing the dimension. The cavitation detection method is shown in Figure 2.7.

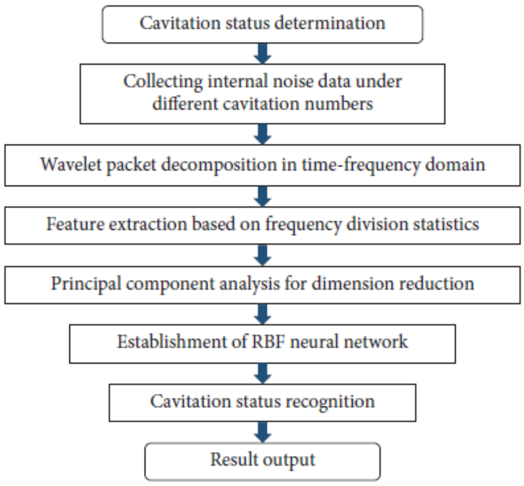

Figure 2.7: Flowchart of cavitation detection by using RBF neural network with WPD analysis and PCA. (Dong et al., 2019)

Principal Component Analysis (PCA) is a statistical method that uses orthogonal transformation to transform datasets of the possibly correlated variables into a set of data which are linearly uncorrelated variables known as principal components (Li & Liu, 2018). The data dimensionality can be reduced, and abnormalities can be removed from the data. RBF is a type of Artificial Neural Network (ANN) that uses radial basis function as the activation function. It is applicable for classification problems with abrupt changes in signals.

Other than that, support vector machine (SVM) algorithm and discrete wavelet transform are used for diagnosis in condition monitoring of the centrifugal pump as well. According to (Ebrahimi & Javidan, 2017), SVM and wavelet transform are used to detect the fault seal, faulty impeller and cavitation of the centrifugal pump, where 44 statistical features are obtained from each of the vibrational signals. The correlation-based feature selection technique in Weka software is used to select superior features as input to the classifier. SVM has various parameters that can be tuned, such as penalty parameter, type of kernel function, kernel parameter and polynomial degree. The authors concluded that accuracy of the classifier increased as the degree of kernel increased or by reducing the width of the RBF kernel. The result of the research shows that the proposed approach is suitable for fault detection of centrifugal pump.

Besides, (Xue, Li, Wang, & Chen, 2014) also uses intelligent diagnosis methods to diagnose frequent faults occurring in a centrifugal pump, such as cavitation, impeller imbalance and shaft misalignment. They reported that a sequential fuzzy diagnosis model was developed by using the possibility functions. Figure 2.8 shows the statistic filter, support vector machine (SVM), possibility theory and Dempster-Shafer theory (DST) are utilised for the vibration signal analysis collected from the centrifugal.

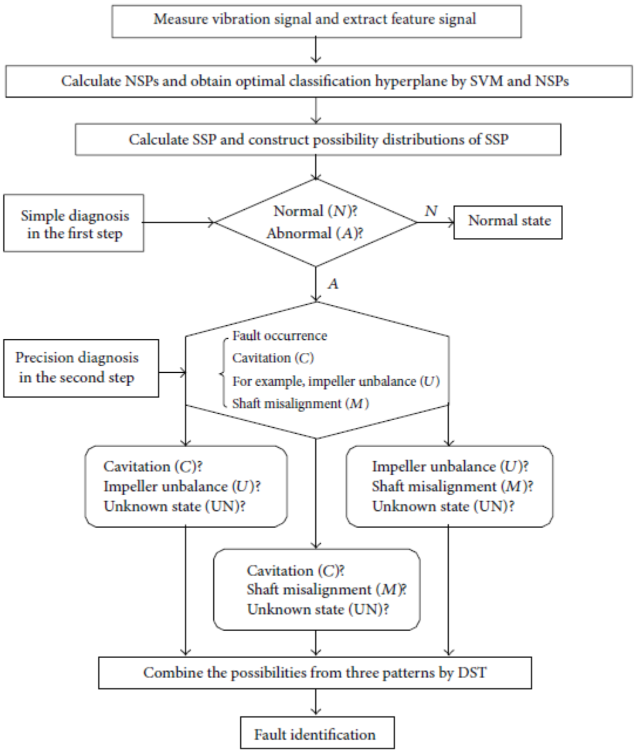

Figure 2.8: Flowchart of sequential fuzzy diagnosis.
NSPs: Non-dimensional symptom parameters, SSP: Synthetic symptom parameter.
(Xue et al., 2014)

Other than that, there are also other machine learning algorithm employed to determine the faults occurred in a centrifugal pump. Multilayer Perceptron (MLP) and support vector machine (SVM) has been used in the oil and gas industry for fault prediction of centrifugal pump. KNIME platform is used to implement the machine learning algorithm. Eight different sensors are used in this research to measure flow rate, bearing vibration, axial displacement and temperature of the motor coil. SVM shows higher precision than MLP, while MLP shows better classification results (Orrù et al., 2020).

(Panda et al., 2018) also use support vector machine (SVM) to predict flow blockages and impending cavitation of centrifugal pumps. Machine Fault Simulator (MFS) is used for experimental purposes. The authors state that the flow blockage in a pump will be very likely causing cavitation. Therefore, early detection of the flow blockage is significant. The marking on the valve shows in the figure below. SVM parameters are being optimized including penalty parameter, RBF kernel hyperparameter and the ratio of training and validation data. Binary classification gives a better result as compared to multiclass classification.

Support vector machine (SVM) is frequently used as the machine learning model to detect cavitation in a centrifugal pump. Figure 2.9 depicted the machine learning process of the research. The authors (Rajesh & Anil, 2017) state that the genetic algorithm (GA) can help to determine the optimal parameters of SVM. Genetic algorithm is an optimization process that uses crossover, mutation and selection operation. Besides, the receiver operating characteristics (ROC) graph evaluates the performance and shows the proposed method is dependable. Based on the experimental result, the authors suggest that this diagnosis method can be applied in reality without any human intervention.

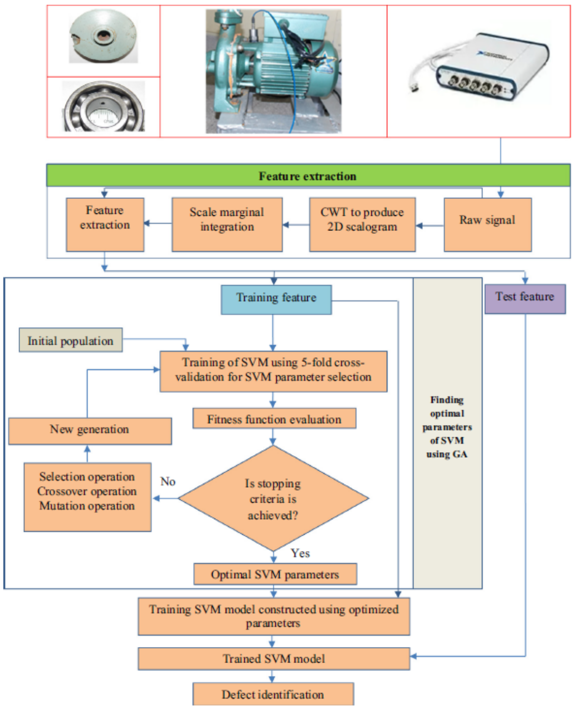

Figure 2.9: Flowchart of automatic fault detection by using SVM and GA. (Rajesh & Anil, 2017)

2.3.2 Acoustic and Vibration Signal

Based on the research, (Al-Obaidi, 2020) used three different approaches for cavitation detection such as vibration technique that uses accelerometer sensor, acoustic technique using microphone sensor and the determination of pressure at suction and discharge using pressure transducer. The Fast Fourier transform technique is used to transform the time domain signals into the frequency domain, where the frequency elements of the signals can be used to represent the mechanical condition of the machine components.

The author claims that the levels of vibration and acoustic amplitudes are lower when the centrifugal pump operates in a low flow rate range as compared to a high flow rate. Flowrates with high vibration amplitudes that correspond to the NPSH performance graph are considered with cavitation. Furthermore, Cavitation Detection Index technique is implemented where the features are normalised by dividing actual values of statistical features to the maximum values in time domain analysis (TDA).

Based on the research result, the author states that the vibration signal is more sensitive as compared to the acoustic signal in determining the development of cavitation. The acoustic technique is incapable of capturing entire changes within the pump, whereas the vibration technique can capture small changes. The normalized values of mean results for vibration and acoustic as threshold can be used to detect cavitation. Lastly, the author concluded that the sensor with a low-frequency range is preferred to determine cavitation in a centrifugal pump as it has better sensitivity and lower cost.

In addition, machine learning using artificial neural network (ANN) for fault diagnosis using vibration signals has been experimented as well. Multilayer perceptron (MLP) which is a feedforward network is used in this case. The time domain features of vibration signals achieved 100% accuracy compared to the frequency domain feature. Less computation cost is required for a time domain-based machine learning model. The input vector that includes data acquired for all four measuring points has a better result than the input vector that process data from each bearing independently (Sepulveda & Sinha, 2018).

Ensemble deep contractive auto-encoder (EDCAE) can handle data collection under noisy environment. Deep contractive auto-encode (DCAE) is designed to learn invariant feature representation automatically. The model can manage noisy data effectively due to the Jacobian penalty term in DCAE. The combination of DCAE, fisher discriminant analysis and Softmax classifier able to produce accurate diagnosis result. It shows that EDCAE is more effective in coping with noisy data than other DAE methods (Zhang, Li, Gao, Chen, & Li, 2020). In this review, EDCAE method shown to be capable of managing data collection for intelligent fault diagnosis of machine even in a noisy environment. As some of the environmental noise could not be avoided due to the location of a pump in a noisy ambience.

The majority of the research works uses vibration data of the centrifugal pump for flow blockage and cavitation detection. Acoustic data is rarely discussed in machinery intelligent diagnosis.

2.3.2(a) Statistical Features

(Panda et al., 2018) reported their ML implementation where statistical features such as the mean, standard deviation, skewness, kurtosis, crest factor and entropy are extracted from the time domain of the vibration signals. Standard deviation is concluded as the best statistical feature.

The discriminant feature extraction method is implemented to solve the problem where raw vibration signals are not sensitive to incipient faults or serious faults. In addition, the discriminant features can be extracted by using three phases.

Healthy baseline signals are selected in the first phase. Then, cross-correlated to the vibration signals of centrifugal pump of different classes. The correlation sequence yields a set of new features. Lastly, the raw hybrid features in time, frequency and time-frequency domain are extracted from healthy and different classes vibration signals in the third phase (Ahmad, Rai, Maliuk, & Kim, 2020).

Based on MATLAB documentation (Mathworks, 2019), the useful statistical features for time domain is mean, skewness, kurtosis, standard deviation, root mean square, etc. In addition, the frequency domain condition indicators can be identified from power bandwidth, peak value, harmonics, mean frequency, peak frequency, etc. The condition indicators identified can be used for fault classification or prediction of the remaining useful life (RUL). However, frequency domain features are able to distinguish the source of vibration as each rotating component rotates at different frequencies. Furthermore, boxplot may be used to learn more about each feature, as the boxes of the classes do not overlap with each other, that feature can easily distinguish the classes.

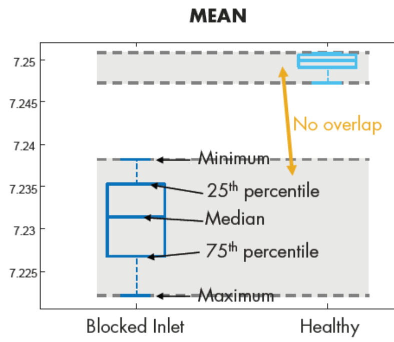

Figure 2.10: Box plot. (Mathworks, 2019)

2.3.2(b) Order Analysis

From the literature survey, there are limited reports on ML-enabled diagnosis of cavitation of the centrifugal pump by using data from the order analysis. Hence, in this work, we are going to discuss the features indicator found in order analysis.

Order analysis is a technique for analysing acoustic and vibration signals in rotating or reciprocating machineries such as engines, compressors, turbines, and pumps. The traditional acoustic and vibration analysis such as Fast Fourier Transform (FFT) is unable to detect the mechanical characteristics that change with speed. Order analysis can be used to identify the source of unwanted vibration by determining the order. First-order refers to the rotational speed of the machine, where second-order is twice the rotational speed. Order analysis is practical for machine condition monitoring (MCM) and acoustic, vibration and harshness (NVH) testing (Sound, Measurement, & Version, n.d.).

There are three types of order analysis which are joint time frequency spectrum, order analysis frequency spectrum and order analysis order spectrum. They are to inspect the development of frequency content of FFT over time, to inspect vibration magnitudes changes with rotational speed and to identify relative overtone amplitudes of rotational speed as the rotational speed increases respectively (NI, 2012).

The behaviours of the centrifugal pump can be discovered from the characteristics found in order analysis. However, the features from order analysis to be used for machine learning are yet to be discovered.

2.4 Machine Learning Algorithm

Based on the literature survey above, the majority of the authors seemingly prefer Support Vector Machine (SVM) as a machine learning model for the fault identification of centrifugal pump. SVM can produce high accuracy in machine condition monitoring and fault diagnosis classification as it is excellent in generalization (Widodo & Yang, 2007).

SVM is commonly used for the classification of two classes of data. However, it can also be done for higher-dimensional space with multiple attributes. The hyper-plane is used as a decision boundary in higher-dimensional space that distinct between classes. A more significant distance between the surrounding data points and the hyper-plane indicates a satisfactory classification. Alternatively, the larger the margin, the lower the generalization error of the classifier (Bordoloi & Tiwari, 2017). Hyperparameters of SVM are C penalty and gamma, where they play a vital role to build a robust and high accuracy model. SVM also can perform non-linear classification with kernel functions. The kernel functions can prevent overfitting of the model, which may occur due to the model's high dimensionality. There are several kernel functions such as linear, polynomial, Gaussian RBF. (Widodo & Yang, 2007)

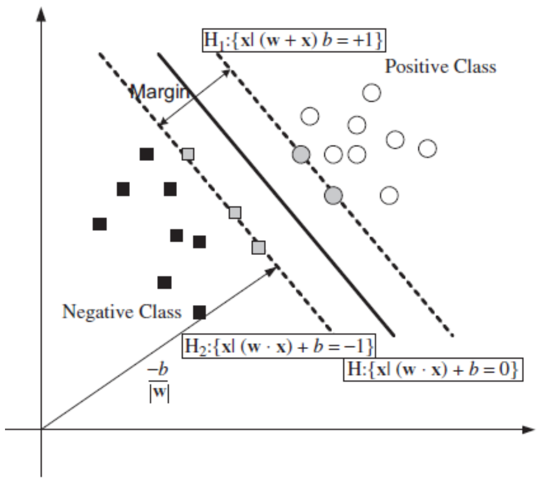

Figure 2.11: Classification of two classes with SVM which shows a linear boundary. (Widodo & Yang, 2007)

MATLAB Classification Learner from the Statistics and Machine Learning toolbox is able to train models with various classifiers for classification purposes. Validation scheme, features selection, classifier type, etc can be specified based on preferences. It can assist a user in choosing the optimizable machine learning model by tuning the hyperparameters automatically with Bayesian optimization. It minimises the model loss according to the selected validation scheme. Automated training for various classifiers can be done to find the best classification model (Mathworks, 2016). The classifiers available in the Classification Learner are decision trees, discriminant analysis, logistic regression classifier, Naïve Bayes classifier, Support Vector Machine (SVM) and Nearest Neighbour. Besides, there are numerous methods to analyse the classification result such as confusion matrix, parallel coordinate, Receiver Operating Characteristic (ROC) curve scatter plot, etc.

Majority of the features mentioned above will be utilised in this project to figure out the most suitable machine learning algorithm for the dataset. Other than that, the Classification Learner application is a user-friendly Graphical User Interface (GUI) that allows us to further analyse each algorithm and to compare the results among various classifiers.

CHAPTER 3

METHODOLOGY

3.1 Experimental Setup

Figure 3.1 below shows the experimental rig available in the laboratory. Sea Pump DTM-20 centrifugal pump is used to pump water from the tank. Sensors used in measuring physical parameters to monitor the flow blockage condition in the centrifugal pump are shown below.

- Accelerometers (Dytran 3055D2T) are calibrated by using a vibration calibrator (Bruel & Kjaer Calibration Exciter Types 4294). Two accelerometers are used in the experimental setup which is mounted on the radial axis (x-axis) and vertical axis (z-axis) of the impeller housing, as depicted in Figure 3.2. Both sensors were mounted perpendicularly. The sensitivity of the accelerometers mounted on the x-axis and z-axis is 102.5mV/g and 101.7mV/g, respectively. They could operate in the temperature range of -55°C to 107.22°C. It was used to collect vibration data generated by the centrifugal pump.
- Microphone (G.R.A.S. 46AD 1/2” CCP Pressure Standard Microphone Set) is calibrated by sound calibrator (G.R.A.S. 42AG Multifunction Sound Calibrator). The microphone was deployed to obtain the acoustic signals generated by the centrifugal pump. The sensitivity of the microphone at 250Hz is 50mV/Pa. The microphone can be operated and stored within the range of -30°C to 85°C and -40°C to 85°C respectively.
- Tachometer (Optical Speed sensor 152G7) is used to measure the rotational speed of the centrifugal pump. A near-infrared light beam will be emitted from the tachometer to the optically reflective strips pasted on the rotational shaft as shown in Figure 3.3. One light pulse will be returned for every shaft rotation.

Next, the data from accelerometers, microphone, tachometer were processed by LMS SCADAS Mobile data acquisition system. LMS Test Xpress software was deployed to control the data recording, settings of the sensors and visualize the data collected. The calibrated value of accelerometers that mounted on the x-axis ad z-axis is 104.221mV and 104.11mV respectively. Besides, the calibrated value of the microphone is 50.4365mV. The values are inserted into the respective channels in LMS SCADAS Test Xpress. The data were collected and sampled at 51200Hz in the data acquisition system and exported as MATLAB data file to further process in MATLAB.

Figure 3.1: Experimental rig of centrifugal pump.

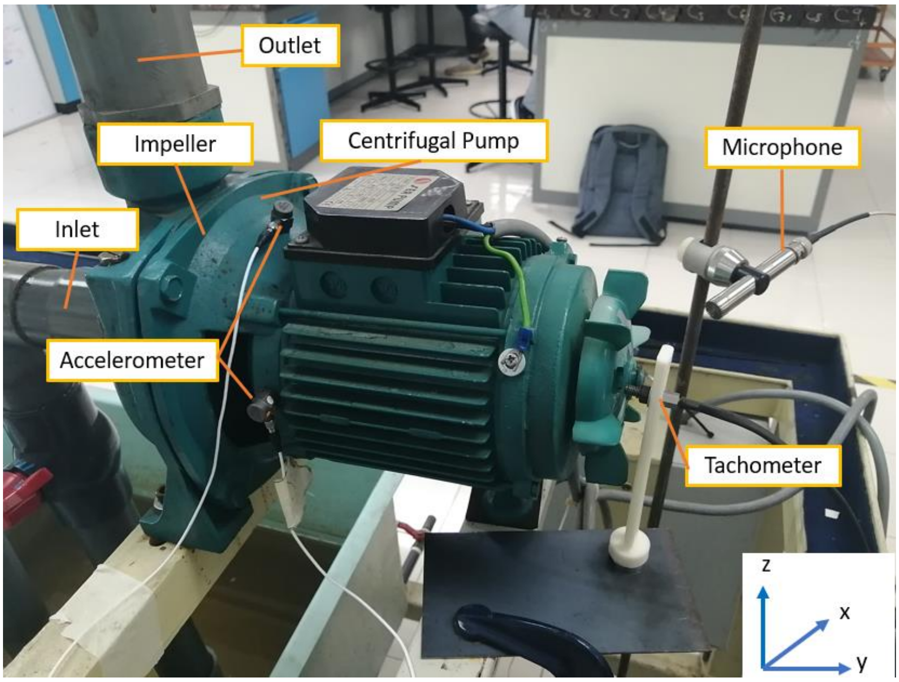

Figure 3.2: Experimental setup of sensors, data acquisition systems and centrifugal pump.

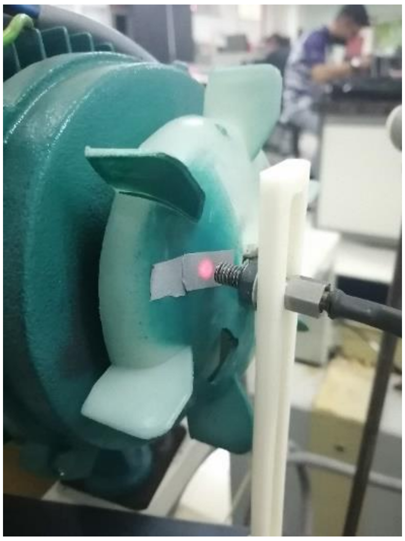

Figure 3.3: Operation of optical speed sensor where near infrared light is emitted to a reflected strip on the shaft of centrifugal pump.

3.2 Raw Data Collection

Data are collected for six different blockage conditions, as shown in Table 3.1. The blockage levels are simulated by adjusting the pump inlet valve according to the marking on the knob. The marking on the knob indicates 20% (inception), 40% (development), 60% (mild), 80% (severe) closure of the inlet valve. These also indicate various blockage conditions of the centrifugal pump. Other than that, one of the bolts that helped to secure the centrifugal pump on the platform is loosened to create instability, as shown in Figure 3.5.

The experiment for each blockage condition is repeated for four times to gain more reliable and consistent data sets and comprehend the prominent characteristic for each condition. As part of the requirement for order analysis, the rotational speed of the impeller blade is adjusted from lowest to highest in each experimental test. The purpose of order analysis is to evaluate the factors that contribute to the vibration of the centrifugal pump.

Table 3.1: Various blockage conditions in centrifugal pump.

Blockage Conditions	**Description**
No blockage	Pump inlet valve is fully open.
Inception	Closure of pump inlet valve is adjusted to 20%.
Development	Closure of pump inlet valve is adjusted to 40%.
Mild	Closure of pump inlet valve is adjusted to 60%.
Severe	Closure of pump inlet valve is adjusted to 80%.
Mild and unstable	Closure of pump inlet valve is adjusted to 60%. One of the bolts that secure the centrifugal pump on the platform is loosen.

Figure 3.4: Knob to adjust opening of inlet valve.

Figure 3.5: Position of bolt being loosen.

3.3 Data Preprocessing

The collected data were preprocessed and visualized by MATLAB R2020b. Machine learning requires a large amount of data to ensure the robustness of the trained model. However, the data collection process might be time-consuming, expensive or even difficult due to the location of the centrifugal pump (Zhou, 2016). Therefore, sample downsizing is implemented to produce slightly more datasets from the collected raw data. Each dataset collected was downsizing for three times with a factor of 5, 10 and 15. Hence there is a total of 72 datasets prepared for the machine learning process. Aside from that, the statistical features of vibration and acoustic data in the time domain and frequency domain are extracted as shown in the following sections. Statistical features extraction is the least time-consuming as compared to other feature extraction methods. The features could be ranked based on the Chi-Square test (Shukla, Yadav, Sharma, & Khare, 2016).

3.3.1 Statistical Features for Time Domain

The raw data of acoustic and vibration signals are collected in the time domain, as analyzing the signals in time domain is fast and straightforward. The chosen features or vibration signals from time domain analysis are simple mathematical equation that requires low computational time. The statistical features extracted from time domain signals and frequency domain signals are referred to a few research papers regarding the vibration and acoustic signals of rotational elements (Caesarendra & Tjahjowidodo, 2017).

- Mean

Mean is the average value of acoustic and vibration data, indicating the 'central value' of a data set. It can help to notice the change of the vibration and acoustic signals of different classes.

$$Mean = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1}$$

- Variance

Variance is the measure of data spread out from the mean value.

$$Variance = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \tag{2}$$

- Kurtosis

Kurtosis is the measure of how the data distribution tail differs from the tail of the normal distribution.

$$kurtosis = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2)^2} \tag{3}$$

- Root Mean Square

Root mean square is commonly used in signal processing analysis as it indicates the energy level of the signals.

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N}|x_i|^2} \tag{4}$$

- Skewness

Skewness examines the asymmetrical spread of the acoustic and vibration signals about its mean.

$$Skewness = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2})^3} \tag{5}$$

3.3.2 Statistical Features for Frequency Domain

Dataset is transformed to the frequency domain by using Fast Fourier Transform for further analysis. Hence the statistical features are extracted from the frequency domain to assist in comprehending the characteristics possessed by each class.

- Mean Frequency

  It estimated the mean of normalised frequency of the time domain signals.

- Maximum

  It identifies the maximum amplitude of the signals in the frequency domain.

- Variance

  It measures how the data spread out from the mean value on the frequency domain.

$$Variance = \frac{\sum_{i=1}^{n}(y - \bar{\bar{y}})^2}{n-1} \tag{6}$$