Feature Extraction and ML Classifier Development FDD

Feature Extraction and Machine Learning Classifier Development for Fault Detection and Diagnosis

By DOYI JOO, B.ENG

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree of Master of Applied Science

MASTER OF APPLIED SCIENCE (2020) McMaster University Mechanical Engineering Hamilton, Ontario, Canada

TITLE:	Feature Extraction and Machine Learning Classifier Development for Fault Detection and Diagnosis
AUTHOR:	Doyi Joo, B.Eng
SUPERVISOR:	Professor Saeid Habibi
NUMBER OF PAGES:	xvii, 139

Abstract

In this research, a fault detection and diagnosis strategy for internal combustion engine is developed using measurements that are readily available in engine testing environment to monitor abnormal combustion. The FDD strategy is designed to monitor the engine on a cycle-by-cycle basis using measurements that are accessible on a running vehicle. Pressure measurements are easily accessible in a testing facility that provide useful insight into the quality of the combustion occurring inside the engine. However, due to its cost and complex installation procedures, it is not feasible to obtain in-cylinder pressure measurements from an in-vehicle engine. Faults of a mechanical system are often investigated using vibration. Due to the low cost and non-invasive nature of accelerometers, vibration measurement is used to monitor the in-vehicle engine. However, as vibration behaviors of complex system such as an engine is hard to characterize, in-cylinder pressure measurement is used during the development of the FDD strategy to assist in characterizing the vibration measurement. Upon data acquisition, features are extracted from the vibration measurements using Extended-MSPCA for better characterization and data reduction with a multi-baseline technique. Pressure measurements are analyzed using thermodynamic theories to assess the combustion quality of each cycle. The vibration measurements are labelled corresponding to the pressure analysis. An artificial neural network classifier is developed using the extracted and labelled features. Developed classifier detected the fault and its location with an overall accuracy of 96.3%.

To my parents, my sister & R.T for their endless love and support

Acknowledgements

I would like to thank my supervisor, Dr. Habibi for his guidance throughout this research. I would also like to thank Dr. Tjong and his team at Ford Powertrain Engineering Research and Development Centre for their support during my time at PERDC.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ADC	Analog to Digital Converter
ANN	Artificial Neural Network
BDC	Bottom Dead Center
CAD	Crank Angle Domain
CMHT	Center for Mechatronics and Hybrid Technologies
CWT	Continuous Wavelet Transform
DWT	Discrete Wavelet Transform
FDD	Fault Detection and Diagnosis
ICE	Internal Combustion Engine
IEMSPCA	Industrial Extended Multi-Scale Principal Components Analysis
MSPCA	Multi-Scale Principal Component Analysis
OBD	On-Board Diagnostics
PCA	Principal Components Analysis
PERDC	Powertrain Engineering Research and Development Center
RBC	Reconstruction Based Charts
RMS	Root Mean Square
SPE	Squared Prediction Error
STFT	Short Time Fourier Transform
SVD	Singular Value Decomposition
TDC	Top Dead Center
WPT	Wavelet Packet Transform

List of Symbols

$\psi(\cdot)$	Mother Wavelet Function
C_ψ	Wavelet Admissibility Condition Parameter
$X_w(a,b)$	Wavelet Transform
T	Principal Component Scores Matrix
P	Principal Component Loadings Matrix
Λ	Eigenvalue Diagonal Matrix
X^T	Transpose of X
$\mathcal{T}^2$	Hotteling's T-Squared Index
$\mathcal{Q}$	Squared Prediction Error Index
φ	Combined Index
Σ_T	Diagonal Matrix of Principal Components
δ	Upper Limit of SPE Index
τ	Upper Limit of Hotteling's Index
Φ	Upper Limit of the Combined Index
ζ	Chi-Squared Distribution Upper Limit
g_φ	Chi-Squared Distribution Weights
h_φ	Chi-Squared Distribution Degree of Freedom
$COV(\cdot)$	Covariance Function
F_c	Fault Isolation Index

Chapter 1: Introduction

1.1 Overview

Fault detection and diagnosis has been a topic of interest for many years in the industrial sector such as in the oil and gas, the power generation and the automotive industries. The origins of FDD is rooted in the field of control systems and more specifically, in automation technology which was enabled by the rise of computers. An early form of automation is called regulatory control, which allows for low level control actions to be performed without a human operator [1]. More recently, due to great improvements in computational capabilities, the realm of automation has surpassed regulatory control and advanced well into more complex tasks such as condition monitoring (including fault detection and diagnosis) of very complex processes and systems.

Fault Detection and Diagnosis (FDD) has mostly been applied to industrial components and systems to prevent events that may bring forth tremendous amount of financial burden if a malfunction were to occur. However, as service and quality of consumer products become more and more in demand and as sophisticated data analysis methods arise such as machine learning algorithms, FDD technologies are in demand by a variety of industries that aim to provide the best quality for service and product.

Currently, the automotive sector is at the verge of a very important change that may redefine the automotive sector itself and the way consumers perceive transportation and vehicles. The factor that are initiating such monumental changes in the automotive sector is the advancement of computing technology that were mentioned previously. Machine learning algorithms have proven that there is very useful information embedded in data that are being collected by sensors that are currently not being extracted. In the automotive industry large quantities of data are being collected every day. One of the areas that have great potential for extracting useful information through tools like machine learning is in testing, where tremendous amounts of valuable data is being collected on important components of the vehicle such as the engine. In this research, the potential opportunities for technological advancement through the access of unused data or unused data source is explored along with development of advanced data analytics methods.

1.2 Research Motivation

The motivation for this research can be viewed from different perspectives. For example, the condition monitoring or FDD technology for internal combustion engines can be developed and utilized in a variety of different sectors of the automotive industry such as in the manufacturing lines, fleet operations, engine diagnostics, and engine controls.

For a car manufacturer, a high productivity and high-quality assurance can be achieved with the help of FDD technology, which could ensure the detection of faulty engine components. Automated FDD strategy can greatly reduce the number of human operators that are required in quality assurance. Automation of quality assurance would

reduce the time required to assess the quality of the components and increase the accuracy rates of the quality assessment, which allows for human operators to attend to more challenging tasks such as mitigating issues around faulty components discovered. Productivity of the manufacturer and the quality of their products can be improved significantly through an automated FDD strategy.

For a fleet operator, in-vehicle FDD technology would be useful in monitoring the health of the engine. Although On-Board Diagnostics (OBD) exists in every consumer vehicle, it can be argued that there is room for improvement in the sophistication of the data analysis methods to broaden the range of faults and the levels of faults that are monitored.

In an auto-shop or dealership, diagnosis of faults in engines can greatly be expedited with the use of the automated FDD strategy.

For engine controls, an FDD technology that is capable of monitoring the engine's combustion quality in real time would be of great significance as control systems could be adjusted in real time to improve engine emissions and performance.

FDD strategies are being research internationally, including at McMaster's Centre for Mechatronics and Hybrid Technologies (CMHT). This research is a continuation of the FDD research that was conducted by the following researchers at CMHT:

1. S. R. Haqshenas – who introduced CAD-MSPCA for its application to internal combustion engines,

2. M. Ismail – who introduced the Industrial Extended Multi-Scale Principal Components Analysis (IEMSPCA) for its application to alternators and starters using sound and vibration, and
3. A. Doghri – who successfully applied an FDD strategy for in-vehicle application using vibration measurements.

1.3 Research Objectives

The objective of this research is to develop a FDD strategy for internal combustion engines that is intended to monitor the engine on a cycle-by-cycle basis to detect intermittent faults. The development of a FDD strategy involves three main stages as follows, [2]:

1. data collection and pre-processing of time series data,
2. feature extraction of dataset, and
3. classifier development using machine learning algorithms.

This research was performed in collaboration with Ford Powertrain Engineering Research and Development Center (PERDC).

1.4 Thesis Structure

This thesis consists of seven chapters that are organized as follow. Chapter 2 presents the literature review of FDD and feature extraction methods. Chapter 3 present the theory behind the feature extraction and classification strategy that is implement in this research. Chapter 4 presents the experimental set-up, data collection, and data pre-processing

implementations. Chapter 5 presents the implementations of the feature extraction strategy, the results, and the observations. Chapter 6 presents the noteworthy findings of the research and the respective recommendations for future research.

Chapter 2: FDD Strategies for Internal Combustion Engines

2.1 Background

In this section, a review of the internal combustion engine and associated FDD strategies is provided. The objective of this section is not to provide an exhaustive overview of each topic, but only the important factors that influence their interaction and integration.

2.1.1 Internal Combustion Engine

An internal combustion engine functions as a prime mover for a wide range of applications by converting chemical energy that are contained in the fuel into useful mechanical power. The type of internal combustion engine that is most often used in consumer vehicles is the spark-ignition four stroke engine using the Otto Cycle [3]. The Otto cycle has four strokes that consist of intake, compression, power and exhaust. These occur over two full revolutions of the crank shaft [4]. As shown in Figure 1, one rotation of the crank shaft results in linear motion of the piston from top-dead center (TDC) to bottom-dead center (BDC) then back to TDC. This repetitive linear motion, which is also known as reciprocating motion, is achieved by the utilization of a crank that converts rotational motion of the shaft into linear motion of the piston and vice versa.

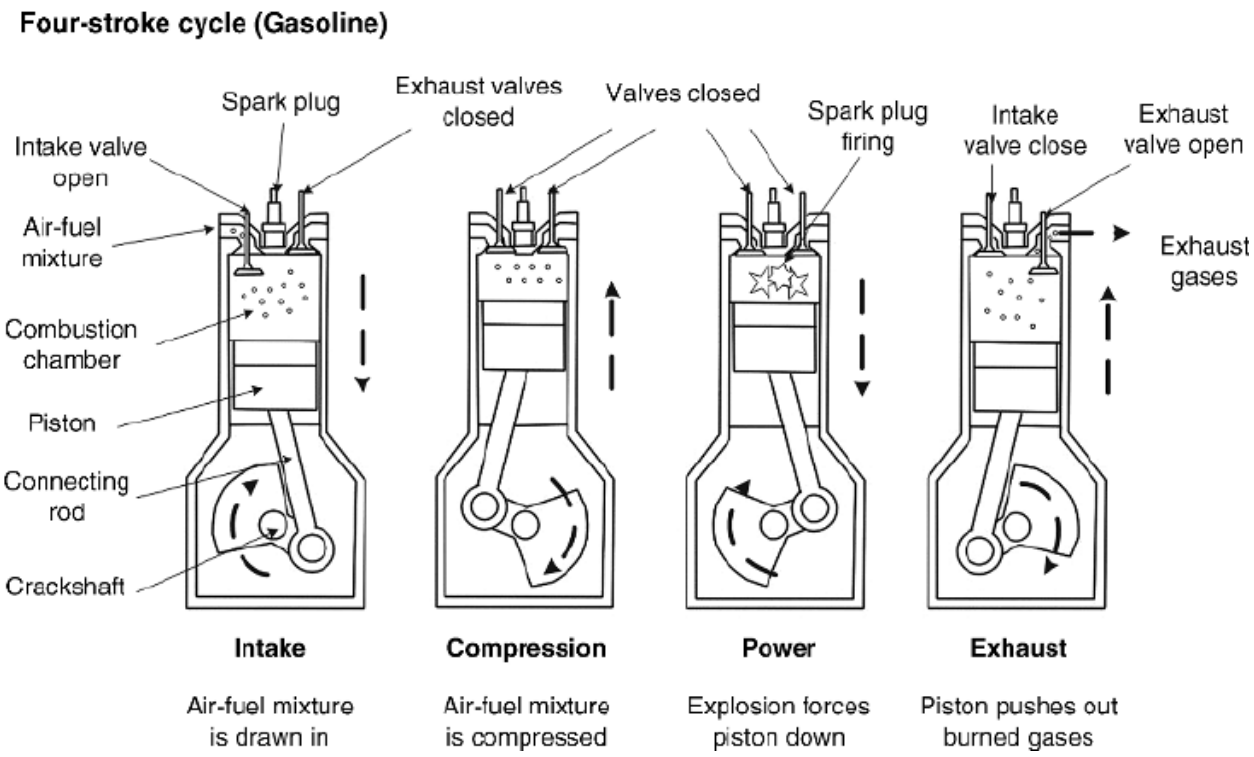


Figure 1: Four-stroke cycle of internal combustion engine with piston and crankshaft position [5]

Each stroke in the Otto cycle can be expressed as a thermodynamic process as shown in Figure 2.

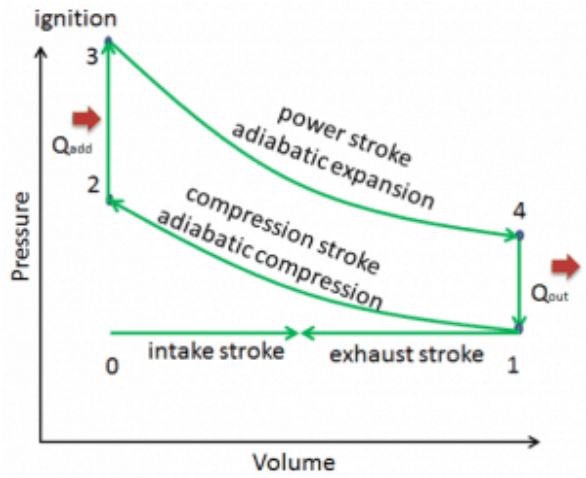


Figure 2: PV diagram of Otto cycle [6]

1. **Intake stroke**: The intake stroke starts with the piston at top-dead center where the volume of the chamber is at its minimum and ends with the piston at (BDC) where the volume of the chamber is at its maximum [3]. As the piston travels down the chamber, the intake valve opens. During this process, the air-fuel mixture enters the chamber while the pressure remains relatively constant [3].
2. **Compression stroke**: The compression stroke begins with the intake valve closing and the piston traveling up the chamber from BDC. This compression is described to be adiabatic and reversible which is also known as isentropic [7]. During this compression, the piston does work on the mixture inside the chamber by decreasing the volume and increasing the pressure inside the chamber [4].
3. **Power stroke**: In between the compression and the power strokes, a thermodynamic process called combustion occurs where the air-fuel mixture is ignited. In the case of a spark ignition engine, which will be the focus of this research, the mixture is ignited by using an electrical discharge. During the very rapid combustion process, the volume stays constant while the pressure and temperature are significantly increased due to the substantial amount of heat that is released inside the chamber. The power stroke, which can be described as an isentropic expansion follows the combustion process [7]. During the power stroke, the piston is pushed from top-dead center to BDC by the gas that is expanded adiabatically upon the combustion of the air-fuel mixture [3]. In the power stroke, pressure decreases while the volume increases and the work is said to be done onto the piston by the gas.

4. **Exhaust stroke**: The exhaust stroke is where the residual gas is removed from the chamber as the piston travels back to top-dead center with the exhaust valve open. The piston movement decrease the volume while the open exhaust valve maintains a constant pressure [8].

The analysis of each stroke shows that the power generated from the cylinder does not occur constantly throughout the cycle instead it occurs in bursts, specifically during the power stroke. This means that a single cylinder engine would deliver power in a pulsating form. Therefore, most engines in a vehicle are equipped with multiple cylinders to provide more consistent power output by staggering the occurrences of the power stroke for each cylinder. Although this results in more consistent power output, it does not eliminate the presence of the pulsation. The fluctuations observed in the multi-cylinder engine occurs more rapidly with smaller amplitude variations. This fluctuation in power delivery is reflected in the rotation of the crankshaft which is never constant in a given engine cycle.

Heywood describes the combustion process that occurs inside the chamber as a fast-exothermic gas-phase reaction [4]. During a normal engine operation, the fuel is mixed together with air and inducted into the chamber through the open intake valve during the intake stroke. As the air-fuel mixture enters the chamber, it gets mixed with the residual gas that remains in the chamber after the exhaust stroke. This mixture of fuel, air, and residual gas from the previous combustion process is then compressed and ignited by an electric discharge of a spark plug. Upon ignition, a flame develops and propagates through the chamber until it reaches the chamber walls where the flame is extinguished [4]. Flames

in a spark-ignition engine are classified as premixed unsteady turbulent flame that propagate in gaseous state [4]. In addition, flame development and propagation vary cycle-by-cycle and cylinder-to-cylinder. And these variations in combustion process is caused by:

1. variations in the mixture motion within the cylinder,
2. variations in the amount of air-fuel mixture in the cylinder, and
3. variations in the ratio of fresh mixture to residual gases in the cylinder [4].

This leads to the fact that no two engine cycles, even for a given cylinder, are identical. Due to this inconsistent nature of the engine, the combustion may not occur optimally for a given cycle since most engines are controlled to run at an "optimal" condition that was configured for an average cycle [4]. Further information regarding internal combustion engines that has not been covered in this section can be found in Heywood's text [4].

2.1.2 Fault Detection & Diagnosis

Fault detection is defined as the capability to recognize that a fault has occurred and fault diagnosis is defined as the ability to locate the fault and the cause of the fault [9]. Such technology was in demand to prevent potentially detrimental failures of machinery, which are arguably the greatest assets of an industrial company, and to prevent extreme costs associated with unexpected machinery or facility downtime. FDD technology poses great benefits as it eliminates the escalation of minor faults through their prompt identification and initiation of timely mitigation measures. Lately, Fault Detection and

Diagnosis (FDD) technology is being pursued by a wide range of industries that are simply looking to provide more reliable products and processes.

Fundamental elements of FDD include detection of malfunctions or abnormalities in the system, isolation and localizing of faults and identification of the fault condition [10]. Before delving into the different concepts of FDD, a few key terms should be explained and defined. A fault can be broadly defined as a state where a measured variable or a calculated parameter within a process or a system is out of its acceptable range [11]. Malfunction of a component can be defined as the root cause for the symptoms exhibited as a fault and leading to failures [11].

Venkakasubramanian suggest that FDD systems commonly follow a series of measurement transformations to arrive at a diagnosis [11]. Figure 3 shows three separate transformations that are involved in determining a fault or failure class given a set of measurements.

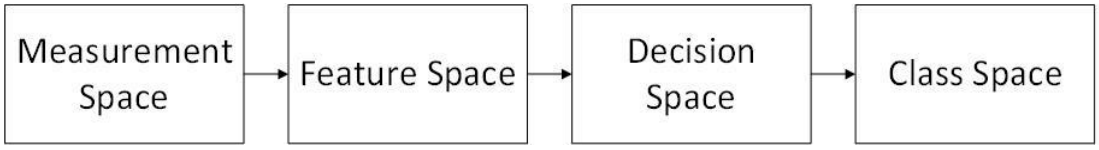


Figure 3: Transformations in FDD systems [11]

The three stage transformations in FDD systems include:

1. Transformation of the input to the overall FDD system called measurements into features through a process called feature extraction. These methods use a priori

knowledge to analyze measurements and obtain useful features, which usually consists of fewer dimensions than measurements [11].

2. Mapping of features to decision through a search or complex learning algorithms without the use of a priori process knowledge. This transformation is commonly performed to minimize misclassification of faults [11].
3. Transformations of the decisions into classes, which indicate different, fault categories including the healthy category. This transformation can be executed using a threshold function accompanied by symbolic logic discriminant [11].

After reviewing the series of transformations that measurements go through in a given FDD system, it becomes evident that feature extraction and diagnostic search strategies are important elements of a FDD system.

When developing a FDD system, there are two different strategies called model-based and signal-based approaches. As the name suggests, the model-based approach utilizes a mathematical model that formulates an expected behavior of the system. The variations in the parameters that results in change of the system behavior is monitored using observers or filters [12]. On the other hand, signal based strategies rely purely on historical measurement data. For complex systems that are challenging to model, signal based FDD may be better suited. This eliminates the need for a mathematical model and the difficulties and limitations of modeling [1]. Instead of mathematical models, signal based approaches attempt to find patterns in the historical measurements. Numerous techniques such as

artificial neural network algorithms and statistical methods are utilized to obtain patterns that is assumed to exist in the data [10].

2.2 Fault Detection & Diagnosis for Internal Combustion Engines

FDD strategies for internal combustion engines have been explored extensively in the past. Earlier FDD strategies for internal combustion engines were motivated by the need for improved productivity and quality in production such presented by Miller and Tjong [13], [14]. These researchers envisioned a system that is capable of detecting manufacturing and assembly defects through an online engine monitoring and diagnostic system [13], [14]. Utilization of such monitoring and diagnostic systems were intended to be integrated in the manufacturing line through a process that was later popularized as cold tests and hot tests [13]. During these tests, the engine is put on a test stand where it is operated in cold or hot modes while measurements are collected at very high sampling rates [15]. Cold test refers to the operation of the engine without combustion and hot test refers to the operation of the engine with combustion. The obtained measurements are analyzed to determine whether or not the engine should proceed to its next manufacturing process [16]. Due to its simplicity and low costs, more research efforts have been found in the development and improvement of cold tests. Following Miller and Tjong, a variety of measurement analysis methods were studied to improve the performance of the cold test procedures [15], [17], [18], [16], [19], [20].

Currently, automation companies are providing fully automated FDD technologies that can be used in the engine manufacturing line. These strategies involve running the

engine in different modes such as cold testing, hot testing with load, and hot testing without load while obtaining measurements such as speed, torque, and pressure. Although it is not clear as to how the measurements are analyzed, it is noted that a variety of defects such as missing piston rings, leaky head gaskets, leaky intake and exhaust valves along with many other defects can be detected using this FDD strategy [21].

More recently, FDD algorithms for in-vehicle applications were explored, where with the appropriate hardware, online monitoring of the engine can be achieved within the vehicle [22]. Currently, vehicles are equipped with On-Board Diagnostic (OBD) systems that monitor the engine closely with a variety of measurements. However, as concepts of autonomous and connected vehicles are becoming closer and closer to reality, it is becoming increasingly important for the automotive industry to broaden the diagnostics for a wider range of components in the vehicle [22], [23]. With ever increasing computation power, more sophisticated methods such as wavelet analysis, Principal component analysis and artificial neural networks (ANN) are becoming tangible for automotive applications [2], [20], [22], [24], [25], [26], [27].

2.2.1 FDD Approach

In the literature, the widely accepted FDD approach for internal combustion engines is the signal-based approach. Model-based strategies have been used but with very limited scope. Wei applied model-based FDD to diesel engines by modeling the in-cylinder pressure and torque produced by each cylinder in crank angle domain based on thermodynamic and dynamics theory [17]. However, in Wei's research, combustive

behaviors were not within the scope of the model as it was intended for the cold test application [17]. Chandroth confirms that signal-based strategies are preferred over the model-based strategies as it is very difficult to accurately model an internal combustion engine, especially the combustive process, given its turbulent and variable nature that was discussed in the previous section [27].

2.2.2 Measurements for FDD

The selection of measurements is very important for signal-based strategies as the correlations that are assumed to exist in the measurements selected act as the a priori knowledge. The type of the measurement such as speed, pressure and/or voltage, the sensor used to obtain the selected type of measurement and the location of the measurement are example of all the decisions that must be made in selecting the measurements for a signal-based strategy. In the following section, the types of measurements that should be considered for the development of FDD for internal combustion engines are discussed.

Vibration Signals

Sound and vibration measurements are commonly used in FDD systems for mechanical systems [28], [29], [30], [31] . Vibration is inherent in any mechanical system that consists of moving components [32]. Every system has its own specific vibration modes that can be observed during its normal operating conditions. These modes are dependent on system parameters such as mass, system stiffness, fitting tolerances and frictions. When changes occur in these system parameters due to a faulty component, observable vibration behaviors may deviate from its normal signature. In addition,

vibration signals can be obtained through non-intrusive methods that are relatively cheap and easy to install. In the literature, a wide range of vibration analysis for FDD of rotating machines are present [1], [28], [31], [33], [34], [35], [36].

In comparison to vibrations of rotating machines, vibrations of internal combustion engines are very complex. Vibration of an engine may arise from a variety of components such as crank train, valve train, piston assembly, fuel system and exhaust system [37]. In addition, the turbulent combustion process contributes to the overall vibration of the engine, which further complicates the task analyzing the vibration. Tjong categorized three main types of faults that may contribute uniquely to vibration:

1. Impulse impact at a specific crank angle due to incorrect clearance between components such as valve train, gears, connecting rod, pistons and bearing;
2. Imbalance in rotating components such as crankshaft, camshaft, balance shaft and gears; and
3. Abnormal combustion [13].

In addition to the complexity that arise from having numerous components, vibration analysis of an engine is especially difficult as vibration contributed by these components are exhibited in different forms. For example, the engine is comprised of components that function in reciprocating motion and rotating motion. Vibration due to rotating components are exhibited harmonically whereas vibrations due to reciprocating motion are exhibited in impulses in the time domain or relative to the crank position.

Furthermore, engine vibrations also comprise of vibration components that are induced by combustion which are often described as stochastic.

Pressure Signals

In the literature, the in-cylinder pressure of an engine is shown to provide a great amount of information on the quality of the combustion in a cycle by cycle basis [4]. Researchers such as Sharkey had performed FDD studies using vibration and in-cylinder pressure signal trace to detect faults such as leaking intake and exhaust valves [30]. Miller, Tjong, Jones and Chandroth presented valuable studies that exhibit great potential for accurate FDD strategies using pressure traces. However, it should be noted that obtaining in-cylinder pressure traces require retrofitting of engine to allow for installation of very expensive pressure transducers.

2.3 Feature Extraction Methods

This section presents a review of a variety of feature extraction methods that are commonly utilized for signal based FDD strategies involving vibrational measurements.

2.3.1 Time Domain Analysis

Time domain analysis is the simplest and cheapest approach to analyzing signals for FDD strategies as measurements are often obtained in time domain. These time series signals are usually referred to as raw signal. A variety of statistical indicators such as crest factor, kurtosis root mean square and standard deviation have been applied to FDD applications for detecting bearing damages [33].

Time domain analysis has been performed on diesel engines for cold test applications, which were aimed at determining the reliable threshold values for pass or fail decision for the cold test given vibrational measurements with very low computational costs [15]. The research concluded that kurtosis coefficients and the RMS value of the vibration signal in time domain was successful at characterizing and detecting faults such as inverted piston, overpressure valve, out of housing exhaust equalizer, improperly tightened oil pump screw and improperly tightened oil jet during cold test operating conditions. Cold test operating conditions refer to conditions at which the engine is operated without combustion. However, time domain analysis is not the most ideal vibration analysis method for FDD strategies for engines at normal operating conditions, which includes combustion, as time domain measurements of engines are very difficult to understand and analysis in time domain yield information in reference to time, which is not very meaningful for the application of FDD on ICE. Time referenced information is not very valuable as engines operate in various different rotating speeds, which would indicate that a fault condition analyzed at one speed would not be able to correlate its learned features at another speed if time domain analysis is applied.

2.3.2 Crank Angle Domain Analysis

As mentioned in the previous section, time domain is not a good reference domain for reciprocating machinery such as internal combustion engine. Instead, measurements obtained in reference to the angle of the crank shaft is widely used. Miller, Tjong and Delvecchio applied statistical indicators from time domain analysis on crank angle domain [13, 15, 14]. Analysis in crank angle domain allows for occurrences of abnormal behaviors

of a signal to be trace to a specific location in an engine cycle which gives insight into to a range of components that may have induced such abnormal behavior. This is because movement of a certain component within the engine is directly related to the position of the crank shaft. Sophisticated methods of converting time domain signal to crank angle domain are explored by Arasaratnam [38]. It is important to note that conversion methods cannot assumed that the crankshaft of the internal combustion engines has a constant rotational speed. If such assumptions were to be made during the conversion process, the reference of the measurements with respect to specific crank angles will be skewed significantly as the rotational speed of an internal combustion engine is not constant. In addition, analysis in crank angle domain eliminates the need for FDD strategy to operate only under the same operating speeds that it was previously trained on. This characteristic is especially important for the development of in-vehicle condition monitoring systems for internal combustion engines.

2.3.3 Frequency Domain Analysis

Frequency domain analysis is a reversible transformation that decomposes the given time-domain signal into their frequency components. These frequency components can be used as features for a given condition. Frequency analysis are most appropriate for stationary signals that are periodic as time information is not available in Fourier transform.

In the literature, frequency analysis such as Fourier transform has been used for FDD of machinery [39], [40], [41]. However, it is often concluded that the event of interest must be periodic for frequency analysis to be effective [42]. Therefore, frequency domain

analysis is not an ideal method to be applied for internal combustion engines where signals are time varying. If internal combustion engines were stationary signals such as in the case of motoring, frequency domain analysis may be able to detect the presence of a fault and narrow down the list of potential culprits. However, if multiples of a component that share the same frequency component exist, the features obtained from frequency domain analysis will not be able to tell which of those components are faulty. This is because the frequency analysis provides information on which frequency appears in the signal but not when the frequency appears in the signal [43].

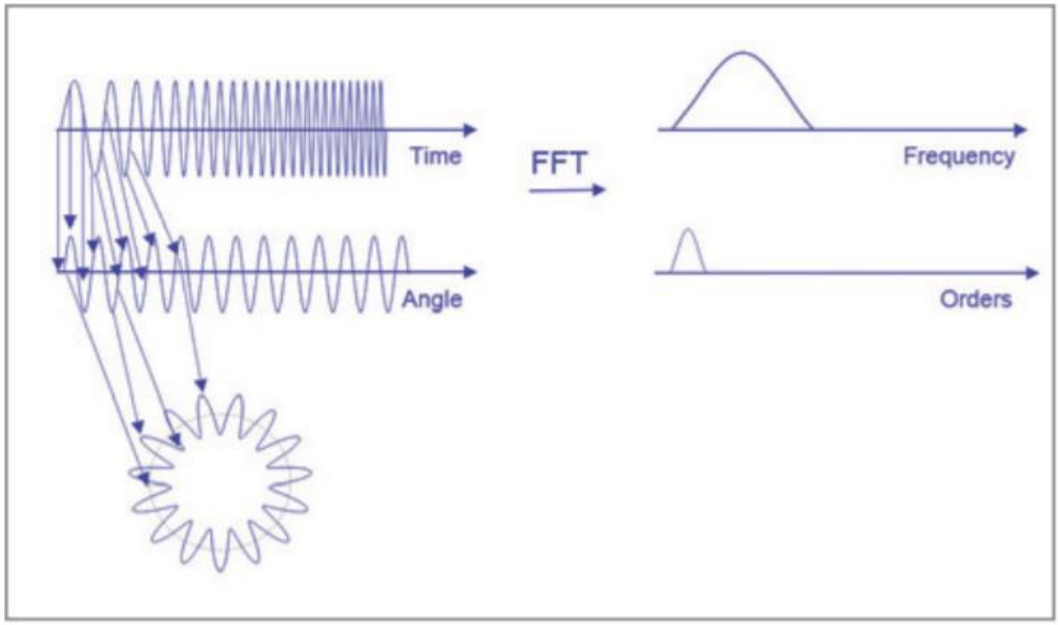


Figure 4: Resampling for order analysis [44]

2.3.4 Order Analysis

Order analysis is a form of frequency domain analysis that is very specific to rotating and reciprocating machineries such as engines, turbines and pumps [45]. Instead

of performing FFT on time-domain signals, order analysis describes a method where FFT is performed on crank angle domain signals. This allows frequency components of non-stationary signals to be extracted in reference to the primary rotating component, which in the case of engines would be the crankshaft. Order analysis provide behavior of harmonic orders of the crankshaft speed with each order corresponding to the multiple of the crankshaft's rotational speed. This method is especially useful for an application where a wide range of speeds are being observed. However, like frequency domain analysis, order analysis is unable to provide information that will aid in identification of a faulty component unless the harmonic order associated to the faulty component is unique in the whole system.

2.3.5 Time-Frequency Analysis

Retrieving the frequency components and the occurrences of those components for a given signal is essential for FDD of rotating and reciprocating machines or for any application that requires the analysis of a non-stationary or transient signal. However, time-domain analysis is only able to provide information on when a component of the signal occurred with respect to time, and frequency analysis is only able to provide information on the frequency components that the signal consists of. To mitigate this constraint and obtain time localization, a variety of approaches that are known as time-frequency analysis was introduced. These approaches include but are not limited to Short-Time-Fourier-Transform, Wavelet Transform, Wigner Transform, and Hilbert-Huang Transform. These approaches map a one-dimensional time-domain signal into a two-dimensional function of

time and frequency. In this literature review, Short Time Fourier Transform (STFT) and Wavelet Transform will be discussed in detail.

2.3.5.1 Short Time Fourier Transform

Short-Time Fourier Transform is a time-frequency analysis method that results from a compromise between time and frequency analysis. STFT achieves the two-dimensional representation of a transient or non-stationary signal in time and frequency through the utilization of windowing technique and Fourier Transform [45]. The windowing technique divides the given signal into short equal time-framed segments, which is individually analyzed using Fourier transform. This windowing technique is applied with the assumption that signals can be considered to be stationary if they are segmented into small enough segments. Figure 5 illustrates the basics of the windowing technique. As shown in Figure 5, the windowing function is first located at time, t = 0, then the window is shifted throughout the signal. The most notable characteristic of STFT is that its window size is kept constant throughout the whole transformation.

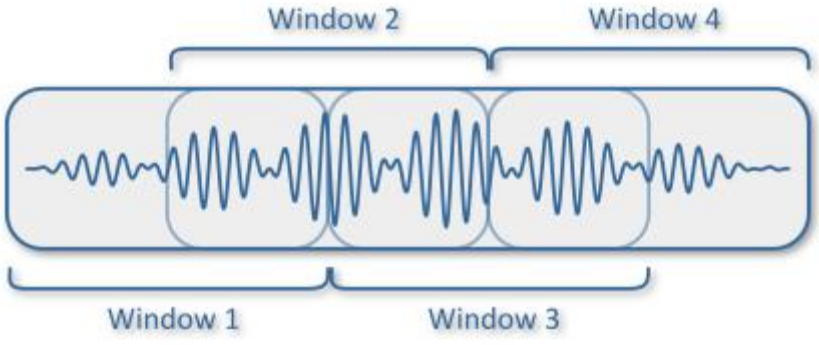


Figure 5: Signal windowing technique [12]

The STFT is defined as:

$$STFT(\tau, \omega) = X(\tau, \omega) = \sum_{t=-\infty}^{\infty} x(t) w(t-\tau) e^{-i\omega t} \tag{2.1}$$

where τ is the time, ω is the frequency, t is the offset, $x(t)$ is the signal to be transformed, and $w(t)$ is the windowing function. And the energy density of the signal at a given time and frequency can be described as:

$$Spectogram(\tau, \omega) = |X(\tau, \omega)|^2 \tag{2.2}$$

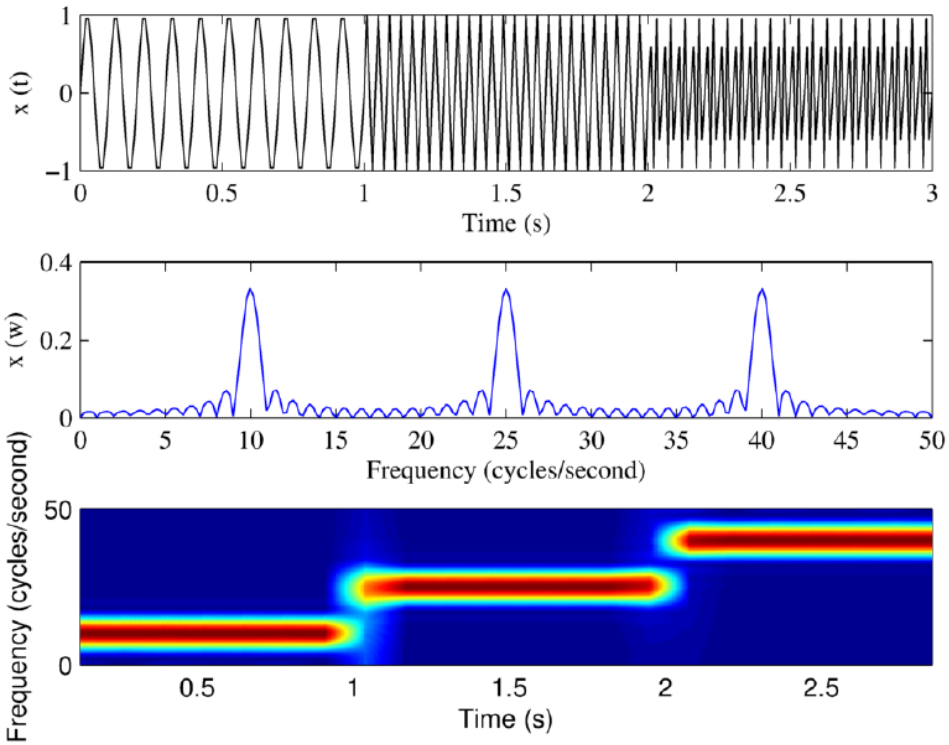


Figure 6: Spectrogram and spectrum of a non-stationary signal [12]

As shown in Figure 6, STFT is capable of simultaneously capturing both the frequency components and its occurrences given a non-stationary signal. However, the compromise between time and frequency domain analysis has its limitation, which is described by the Heisenberg uncertainty Principal. It infers that the increase in resolutions cannot be achieved concurrently due to the inability to know the exact time frequency representation of a signal. Figure 7 and Figure 8 depict the two extreme cases of high frequency and high time resolution. It can be observed that when the resolution of one domain is increased, the resolution of the other decreases. For example, when the size of the window is decreased the time resolution improves while frequency resolution decreases and vice versa when the size of the window is increased. This phenomenon is often referred to as the resolution problem.

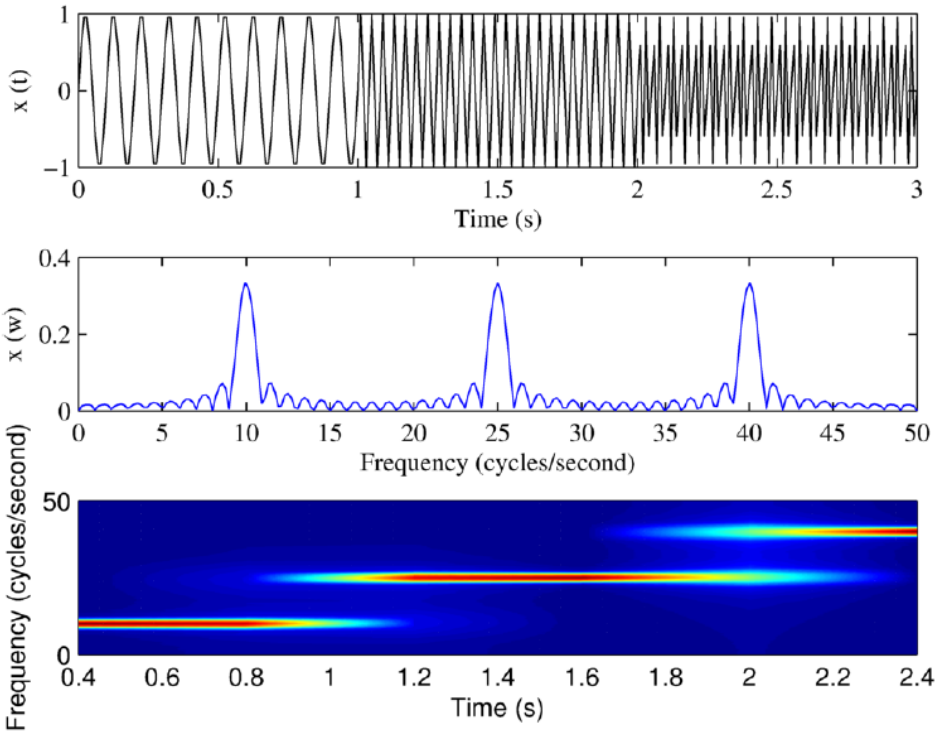


Figure 7: High frequency resolution [12]

A FDD strategy using STFT was applied to bearings by Yazici, where it was reported that high accuracy was achieved [46]. However, the application of STFT on engine measurements may not be enough for the development of an effective FDD strategy for engines. Localization of frequency components in the crank angle is an essential characteristic of the feature extraction methods for FDD of engines due to their non-stationary and transient behavior. STFT has been applied to the analysis of engine vibration by Chen, where different vibration sources are attempted to be identified and characterized. The research concluded that STFT is capable of identifying events such as strong valve impact and piston slap but unable to separate closely-overlapping or weak events such as valve closures of different cylinders. This finding supports the prior judgment that STFT is simply not sophisticated enough for the objective of the FDD strategy that is being explored in this research [50]. Due to the fixed resolution in time and frequency in STFT, other Time-Frequency techniques such as wavelet analysis are suggested in the literature for the analysis of engine vibrations [47]. [48], [49].

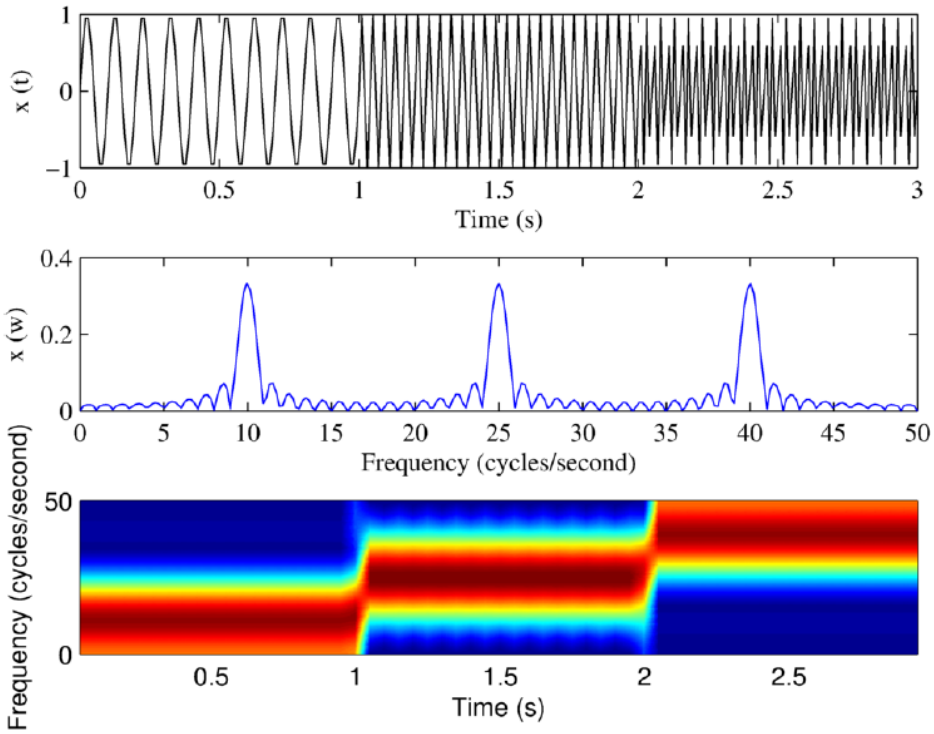


Figure 8: High time resolution [12]

2.3.5.2 Wavelet Transform

Wavelet transform is another form of time-frequency analysis that has been developed as an alternative to STFT that enables analysis of signals consisting of transient, non-stationary and time-varying characteristics. Unlike STFT, where the resolution of the signal is fixed in both time and frequency domains throughout the whole transformation of the signal, wavelet utilizes different window sizes to develop a multi-resolution analysis (MRA) in the time and frequency domain of a signal. In this section, the fundamental review of continuous wavelet transform (CWT), discrete wavelet transform (DWT), and wavelet packet transform (WPT) will be discussed.

The word wavelet originates from the equivalent French word ondelette, which translates to small wave. It is defined as continuous functions that decay to zero at both ends at different rates, which provide compact support in comparison to continuous sinusoids. The rate of decay is altered by the dilation or compression of the wavelet. The wavelet is then translated to capture localized events along with the time domain of the signal that is to be analyzed. These characteristics allow wavelet transform to highlight details of the signal being analyzed, which other forms of analysis may not be able to. Therefore, the ability to detect such minor details of a signal that is very favorable in condition monitoring and FDD application.

The method is established by very simple intuition regarding frequency components. It is known that low frequencies can be observed throughout longer time periods and that higher frequencies can be observed in very short time periods. With this intuition, it was suggested that a large window size be used to obtain information regarding low-frequency components and very small window sizes to obtain information regarding high-frequency components. The ability to capture both the time and frequency components of a signal with high resolution makes wavelet transform a promising method for analyzing transient signals that consist of both high and low-frequency components. It is important to note that the Heisenberg uncertainty Principal still applies to wavelets, however, wavelet transform is a very clever work around method. The resolution problem is mitigated by prioritizing the resolution that is more important at a given frequency range.

The difference between STFT and wavelet analysis is best presented in Figure 9. The wavelet transform's multiresolution property is illustrated by the different sizes of the boxes. The boxes indicate the frequency and time resolution at which each transform can analyze the signal. As illustrated in Figure 9, time and frequency domain analysis are only able to provide information on the signal at their own domain. This is indicated by their rectangular shape that has a fixed width in the frequency domain for time-domain analysis and time domain for frequency domain analysis. STFT is shown to have a square shape with the same width for time and frequency. This infers that STFT has a fixed resolution in both time and frequency throughout the entire frequency ranges. On the other hand, the wavelet transform has both rectangular and square shapes in different sizes. Every box in the wavelet transform has the same area due to the Heisenberg uncertainty Principal; however, the widths of both frequency and time are altered, providing different proportions of resolutions at different frequency levels. For example, at low frequencies, the boxes have larger heights with thinner widths, which indicates that it has better frequency resolution with lower time resolution. And at high frequencies, the boxes are shorter and wider, indicating that it has lower frequency resolution but higher time resolution.

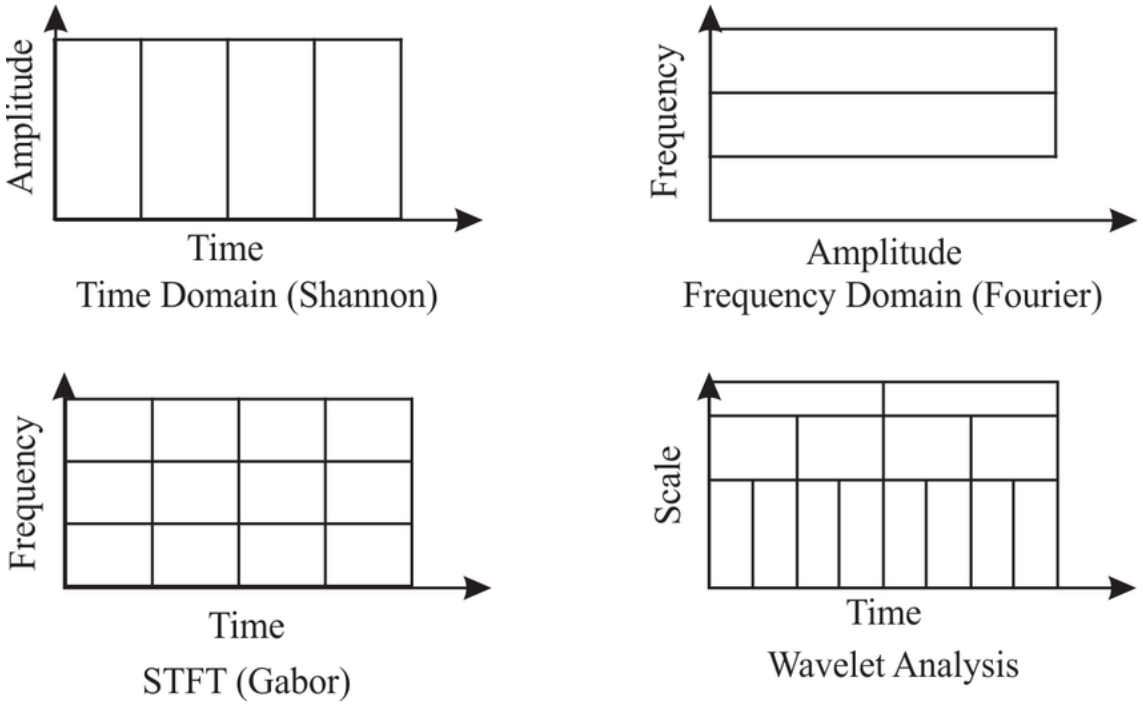


Figure 9: Frequency and Time Resolution of analysis in different domains [46]

Continuous Wavelet Transform

Similar to STFT, wavelet transform multiplies the signal with the wavelet function, which acts as STFT's window function, at different time segments in the signal. The continuous wavelet transform and the inverse continuous wavelet transform are described in Equation 2.3 and 2.4, respectively.

$$X_w(a,b) = \frac{1}{|a|^{1/2}} \int_{-\infty}^{\infty} x(t)\psi\left(\frac{t-b}{a}\right) dt\,, a \in \mathbb{R}^+ \; and \; b \in \mathbb{R} \tag{2.3}$$

$$x(t) = \frac{1}{C_\psi} \iint_{-\infty}^{\infty} X_w(a,b)\, \psi_{a,b}(t) \frac{da\, db}{a^2} \tag{2.4}$$

where a is the scale factor, b is the translational value and $\psi(t)$ is the mother wavelet where children wavelets are derived from given the following equation.

$$\psi_{a,b}(t) = \frac{1}{|a|^{1/2}}\psi\left(\frac{t-b}{a}\right) \tag{2.4}$$

Wavelet analysis is a process where different scaled and time-shifted versions of the mother wavelets are compared with the signal that is being analyzed to measure the similarity between the signal and the wavelet function. The scale parameter corresponds to frequency resolution of the analysis where higher scale provides more global behaviors of the signal and lower scale provide more detailed information that may only last for a short time span. Scaling parameter achieved varied frequency resolution analysis through dilation and compression of the wavelet function [47].

Unlike Fourier transform, the wavelet functions are localized in both time and frequency [47]. Due to this localization, the wavelet functions must be translated throughout the time duration of the signal to ensure characterization of the entire signal. Similarly, wavelets must be scaled in time and shifted in frequency domain due to the localization of wavelets in frequency. For these reasons, the continuous wavelet is described as a process where the signal being analyzed is multiplied with analysis windows that are changed in scale and shifted in time then integrated over time. The square root of the scaling factor is an energy normalization term which ensures that wavelets of different scales have the same amount of energy.

All mother wavelets and its offspring have two main characteristics that are very important for normalizing the transformation. First characteristic is described as,

$$\int_{-\infty}^{\infty} \psi_{a,b}(t) = 0 \tag{2.5}$$

which basically states that all wavelets must have zero average. Second characteristic is described as,

$$\|\psi\| = 1 \tag{2.6}$$

which states that their energy must equal to unity. The ability to obtain the transformation inverse is very useful in Fourier transform. In wavelet transform, dual wavelet function which is a synthesis function for a given wavelet function is utilized to perform the inverse wavelet transformation. To ensure that inverse wavelet transformations are possible, mother wavelets are chosen under admissibility condition which is described in the following equations.

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\psi(\omega)|}{\omega} d\omega < \infty \tag{2.7}$$

The admissibility condition requires that $\psi(\omega)$ is differentiable. There are numerous amounts of wavelets that have been developed for different applications to analyze transient, time-variant, and/or non-stationary characteristics. The selection of wavelet is very important for the success of the application. Although the concept of wavelet transform is very powerful, application of wavelet transform in a continuous form is impractical on digitized computers. In the following section, a discrete form of wavelet that can be applied on computers is introduced.

Discrete Wavelet Transform

Discrete wavelet transform is implemented using multi-resolution analysis method to reduce computation time through elimination of redundancy while still ensuring that an appropriate amount of information is provided for analysis and synthesis of the signal. Multi-resolution analysis method suggests the decomposition of the input signal into two different signals using two channel filter bank and down sampling of the filtered signal. In order to analyze the signal at different scales, the cut-off frequencies of the filter are altered. The signal is essentially being decomposed into different frequency bins using a series of half band high and low pass filters. This process is illustrated in Figure 10 where g[n] and h[n] are filters which are associated with the mother wavelet and scaling function [48]. The following equation describes the relation between the two filters.

$$g[N-1-n] = (-1)^n \cdot h[n] \tag{2.8}$$

The filtering of signals is described by a convolution operation of the signal and the impulse response of the filter, as described in the following equation.

$$x[n] * h[n] = \sum_{k \to -\infty}^{\infty} x[k]h[n-k] \tag{2.9}$$

The low pass filter removes all the frequencies that are higher than the half of the full frequency range of the signal and the result is known as approximation. The high pass filter removes all the frequencies that are lower than the half of the full frequency range of the signal and the result is known as detail. Through this process, half of the information from the original signal is lost which results in the reduction of the resolution in half. The

filtering operation varies the frequency of the signal. The down-sampling process upon filtering alters the scale through the reduction of the sampling rate, which is achieved through the removal of samples. Down-sampling process eliminates redundancies that exist in the filtered signal according to the Nyquist sampling theorem and doubles the scale. The following equations describes how approximate and detail representations are obtained.

$$cA = x_{low}[k] = \sum_{n} x[n] * h[2k - n] \tag{2.10}$$

$$cD = x_{high}[k] = \sum_{n} x[n] * g[2k - n] \tag{2.11}$$

This process of filtering and downsampling to obtain detailed and approximate representations of a given signal can be performed iteratively to break down the original signal into smaller frequency ranges. The decomposition of a signal reduces the time resolution in half as the number of sample has been reduced by half. On the other hand, the decomposition of a signal doubles the frequency resolution as the frequency band of a signal spans only half of the original frequency band [47].

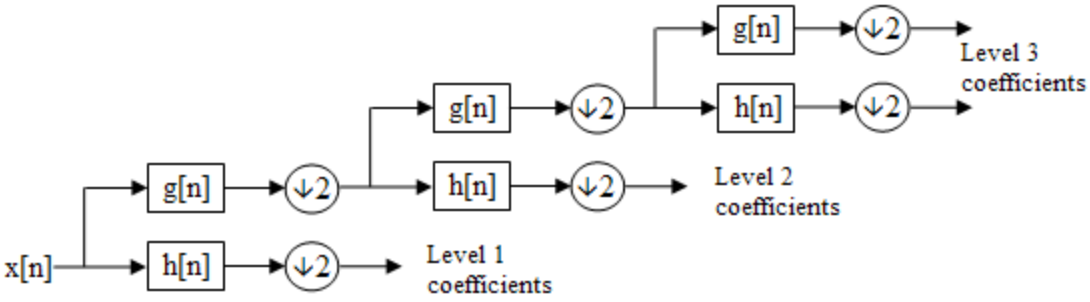


Figure 10: Discrete wavelet transform filter bank [12]

Reconstruction of the original signal can be obtained using Equation 2.12, where y_{high} and y_{low} are given by Equations 2.12a and 2.12b. The summary of the decomposition and synthesis of filter bank is shown in Figure 11.

$$x[n] = \sum \left(y_{high}[k] \cdot g[-n+2k]\right) + \left(y_{low}[k] \cdot h[-n+2k]\right) \quad (2.12)$$

$$y_{high}[k] = \sum_n x[n] \cdot g[-n+2k] \quad (2.12a)$$

$$y_{low}[k] = \sum_n x[n] \cdot h[-n+2k] \quad (2.12b)$$

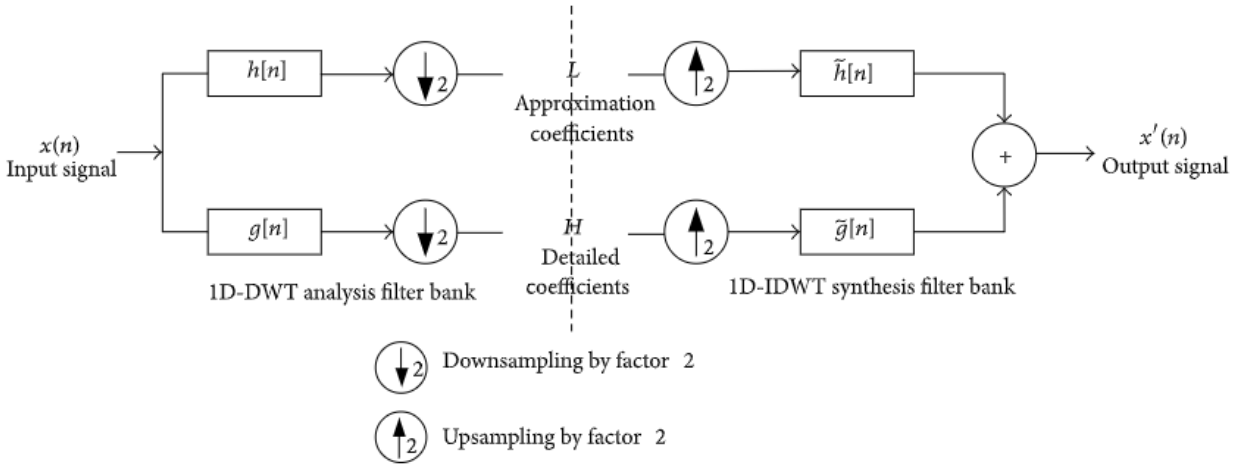


Figure 11: DWT decomposition and synthesis overview [49]

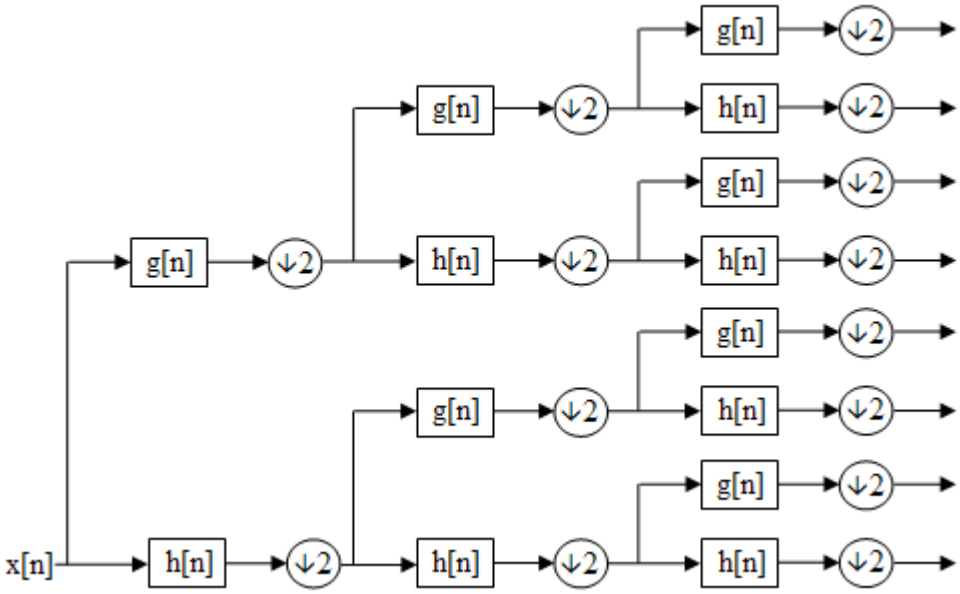


Figure 12: Wavelet Packet Transform filter banks [12]

Wavelet Packet Transform

Wavelet packet transform is very similar to the discrete wavelet transform in that signal is transformed by passing it through a series of high-pass and low-pass filters. However, the wavelet packet transform differs from the discrete wavelet transform as it passes both approximation and detail coefficients through low and high pass filters. Figure 12 shows the wavelet packet decomposition. The main benefit of wavelet packet transform is that the entire frequency spectrum is highlighted equally unlike DWT, which mainly highlights the high frequency components. The difference in distribution of the frequency bands between DWT and WPT can be observed in Figure 13.

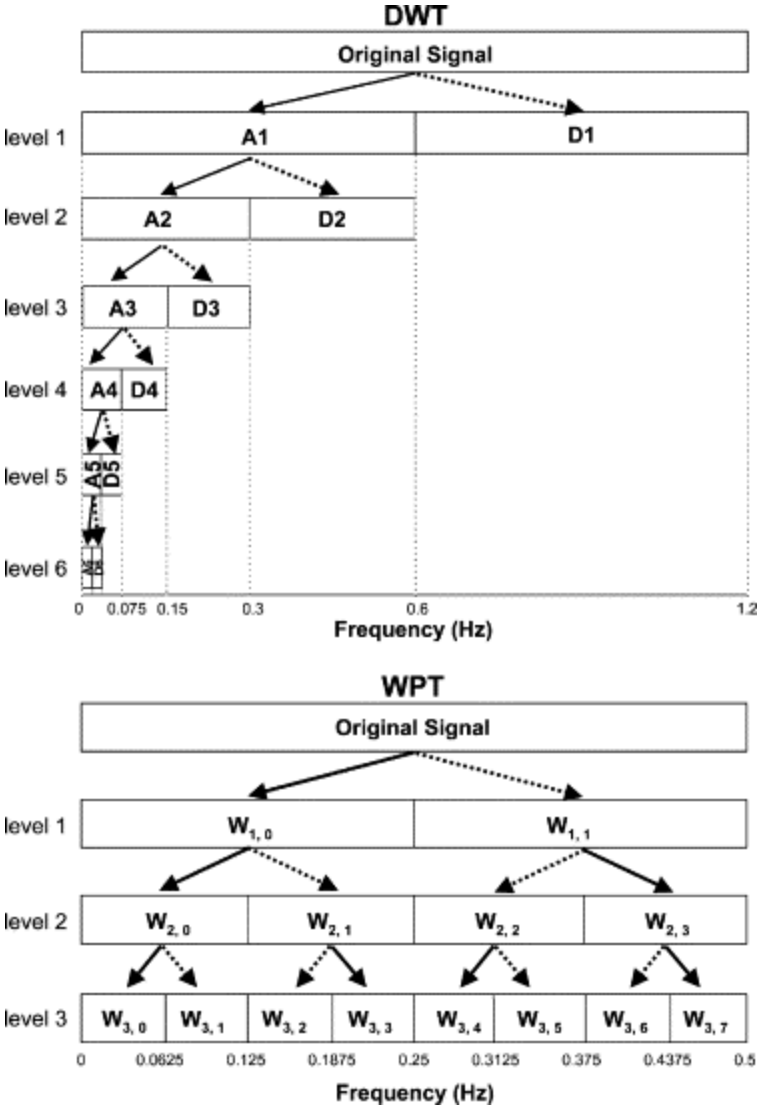


Figure 13: Frequency distributions of DWT and WPT [22]

2.3.6 Principal Component Analysis

Measurements from multiple sensors, either from the same type or different types of sensors, are required as the complexity of the system that is being observed is increased. For fault detection and diagnosis applications, behaviors that relate to faults must be extracted from these multi-sensory measurements using multivariate analysis. Multivariate

analysis refers to statistical methods in which more than one statistical variable is observed and analyzed. Multivariate analysis aims to obtain relationships that exist among the different variables which effect the system's behavior and to obtain dominant variables that contribute the most to the system's behavior through mathematical conversion of measurements into statistical indices. One of the key factors that must be accounted for is the amount of cross-correlation that exists in the entire dataset. The level of cross-correlation would indicate that, although many different measurements were obtained, very little information can be extracted from it. Depending on the situation, an additional measurement may reach a point of diminishing return due to the lack of uniqueness that exist in the supplementary measurement. Principal component analysis is an example of a popular statistical method for multivariate analysis in the field of engineering including fault detection and diagnosis.

Additionally, Principal Component Analysis (PCA) assumes existence of correlations among a set of multiple variables that are provided for a given system. With that assumption, PCA aims to convert the dataset into principal components using orthogonal transformation. In other words, a dataset with many variables that are assumed to be correlated are transformed into a new space composed of principal components, which are a set of linearly uncorrelated orthogonal axes. The contribution of each principal component on the system's behavior is dictated by the variance that exists in the data along the principal component axes. A given principal component's contributions to a system's behavior is high if large variance is observed along its axes. For example, the first principal component has the largest variance of the dataset, the second principal component has the

largest variance of the dataset given the constraint that it must be orthogonal to the first component and so on. This concept is illustrated in Figure 14.

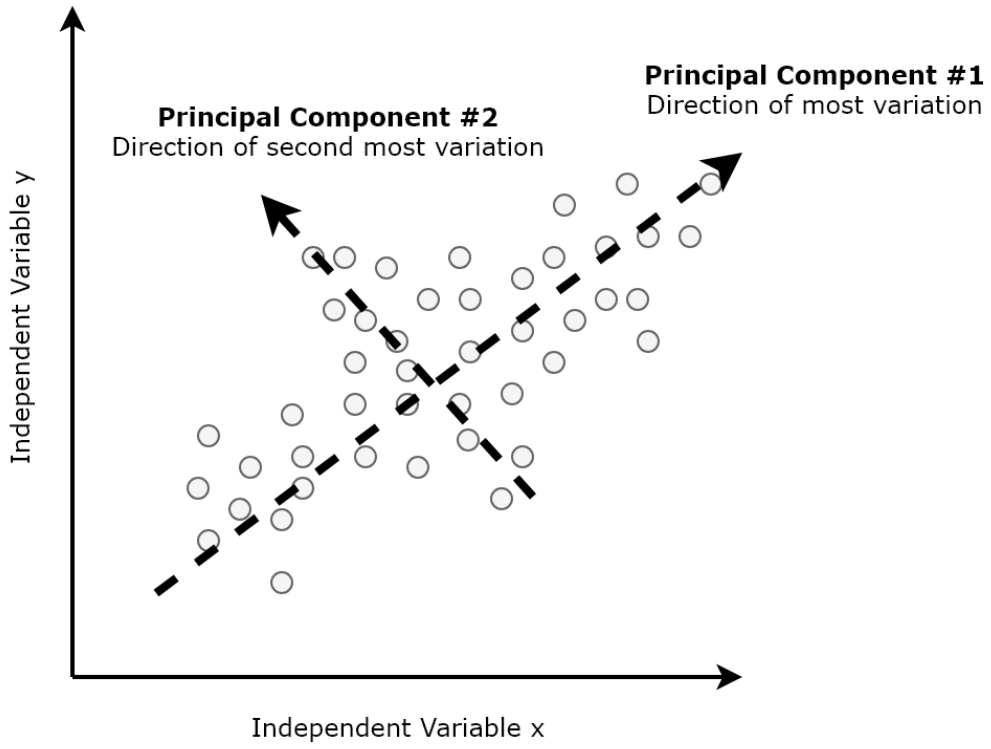


Figure 14: Principal component analysis [50]

During this transformation the number of variables is significantly reduced by eliminating the data points that dependent on principal components with very low variances which are considered to have very little contribution to the system's behavior. This reduction in data set is often referred to as dimensionality reduction. In addition, PCA performs statistical analysis on the compressed dataset to understand the dynamics of the system. To summarize, PCA is a simple and non-parametric form of feature extraction method that extracts the dominant dynamics that exist in the complex dataset, which is

very applicable in the field of fault detection and diagnosis.

$$T = XP \tag{2.12}$$

PCA transform is described by Equation 2.12, where X is the data matrix of size $n\,x\,m$, T is the principal component scores matrix of size $n\,x\,m$, P is the principal component loadings matrix of size $m\,x\,m$, n is the number of measurements, and m is the number of variables. Principal component loadings represent the basis vectors of the new principal component dimension. Principal component scores correspond to the uncorrelated signal representation of the dataset and is organized from highest to low variance components. This indicates that first couple of columns of the principal components scores hold the most information regarding the data set when compared to the last couple of principal components. This implies that PCA can identify which measurement contains the fault signature of interest for the application.

Every column of P represents a basis vector in the new principal component dimension. There are two conditions that must be met by the transformation matrix P. These conditions states that the basis vectors must be orthogonal and that the magnitude of variance influences the principal direction [24]. In addition, the principal component scores matrix must be a diagonal. A diagonal matrix is defined as a matrix whose entries outside of the main diagonal equals to zero. As described by Equation 2.12, the aim of PCA is to transform the data that are described by correlated variables into data that are described by uncorrelated variables. In other words, the main purpose of PCA is to remove cross correlation that exist in the raw dataset. Mathematical, this would translate to obtaining the orthogonal transform matrix P that maps the raw dataset into transformed diagonal

covariance matrix [24]. The derivation of the solution is introduced in the following series of equations with the calculation of covariance of the transformation matrix T, where $A = X^T X$ is a square $m\,x\,m$ symmetric matrix, which can be factorized using the spectral decomposition of a matrix formulation as shown in Equation 2.14.

$$\sum T = \frac{1}{n-1} T^T T$$

$$\sum T = \frac{1}{n-1} (XP)^T XP$$

$$\sum T = \frac{1}{n-1} P^T X^T XP$$

$$\sum T = \frac{1}{n-1} P^T (X^T X) P$$

$$\sum T = \frac{1}{n-1} P^T AP \tag{2.13}$$

$$A = V \Lambda V^{-1} \tag{2.14}$$

Matrix V consists of eigenvectors of A and Λ is a diagonal matrix which consists of eigenvalues that correspond to eigenvectors in V [24]. Due to the symmetric nature of matrix A, orthogonality of its eigenvectors can be inferred. A matrix property which states that a transpose of an orthogonal matrix is equal to its inverse is applied to Equation 2.14 to arrive at Equation 2.15. When Equation 2.15 is substituted into Equations 2.13, Equation 2.16 is obtained and P is set to V to arrive at a diagonalized covariance of the transformation matrix T. Equation 2.17 shows the final form of ΣT.

$$A = V\Lambda V^T \tag{2.15}$$

$$\sum T = \frac{1}{n-1} P^T (V\Lambda V^T) P \tag{2.16}$$

$$\sum T = \frac{1}{n-1} \Lambda \tag{2.17}$$

A factorization method of matrices called Singular Value Decomposition (SVD) theorem is utilized in obtaining the eigenvectors to reduce exposure to numerical errors. Equation 2.18 describes the SVD theorem, where X is any arbitrary matrix with size $n\ x\ m$, U is a unitary matrix of size $n\ x\ n$ called left singular matrix, V is a unitary matrix of size $m\ x\ m$ called right singular matrix, and Σ is a rectangular diagonal matrix of size $n\ x\ m$ called singular values matrix.

$$X = U\Sigma V^T \tag{2.18}$$

SVD is applied to calculate PCA as shown in Equation 2.19 given that U is a unitary matrix and $U^T U = I$. The relation provided in Equation 2.20 relates the SVD and eigenvalue deposition which infers that transformation matrix P is equal to the right singular matrix V since $P = \mathrm{V}$.

$$X^T X = V\Sigma^T U^T U\Sigma V^T$$

$$= V\Sigma^T \Sigma V^T \tag{2.19}$$

$$\Sigma = X^T X \tag{2.20}$$

The application of SVD ensures that columns of V, the eigenvectors and the corresponding elements of Σ, the eigen values are organized in the descending order which means that the first principal components will have the highest variance. The small

eigenvalues are often eliminated in practice as these components of the measurement have small signal-noise-ratio. Therefore, the elimination of negligible eigenvalues may reduce noise or error that exist in the dataset.

To summarize the procedure of PCA transformation is provided in the following:

1. Transformation of the given data set to the principal component dimension by calculating the principal loadings, scores and eigenvalues;
2. Setting of a minimum threshold for the eigenvalues;
3. Elimination of the corresponding loadings and scores of those eigenvalues that do not meet the minimum threshold;
4. Reconstruction of the remaining matrix to original dimension using the inverse PCA transform described in Equation 2.21, where $\hat{T}$ and $\hat{P}$ are the remaining scores and loadings upon elimination.

$$\hat{X} = \hat{T}\hat{P}^T \tag{2.21}$$

The method of defining the threshold limit is a field of its own that is being researched extensively as it is crucial for the performance of PCA. Heuristic rule and Kaiser's rule are two most common methods of thresholding.

The heuristic rule states that components that consists of 95% of the total variance should be considered which infers that any principal component with an eigenvalue higher than 0.05 should remain. The Kaiser's rule states that eigenvalues that are greater than the average eigenvalue must be considered.

In addition to thresholding methods, different pre-processing methods available for the application of PCA should be considered. Non-centered PCA is the method of PCA application with no prior processing of the raw dataset. Centered or covariance PCA is method where the mean value of each variable is subtracted from each instance. And correlation PCA is a method where dataset is scaled and normalized to have zero mean and unit standard deviation by subtracting the mean value and dividing it by the variable's standard deviation. For applications where the variables are in different units or where the variances differ significantly correlation PCA is often utilized.

PCA has been widely utilized in signal based FDD applications due to its ability to quantify the correlation between the components of a multivariate dataset and the system's behavior. This quantification, which provides measure of the covariance, allows for the ability to highlight the components that have the most significant influence on the system's behavior without having in depth technical knowledge or a mathematical model of the system. PCA technique is especially useful in signal based FDD approaches where system is complex, and dataset is composed of high dimensional, noisy and correlated data [51].

For the application of PCA on FDD, Cherry and Qin suggests decomposing the multidimensional data collected for a given system into two smaller sets as shown in Equation 2.21 [52]. $\hat{X}$ represents the main features that consists of high varying principal components and $\tilde{X}$ represents the residuals that consists of low varying principal components. Equations 2.22 shows how the principal component scores and loading

matrixes are decomposed in a similar manner. The method of decomposition is performed through a thresholding technique.

$$X = \hat{X} + \tilde{X} \tag{2.21}$$

$$X = \hat{T}\hat{P} + \tilde{T}\tilde{P} \tag{2.22}$$

$$= [\hat{T}\ \tilde{T}][\hat{P}\ \tilde{P}\ ]$$

$$= TP^T$$

The thresholding is applied to the variance of the entire dataset by setting a criterion on the variance, where meeting the criterion means that the principal component in question belongs to $\hat{X}$ subset and inability to meet the criterion means that the principal component in question belongs to $\tilde{X}$ subset. If p is the total number of principal components, z number of principal components belong to the $\hat{X}$, feature subset while $p - z$ number of principal components belong to the $\tilde{X}$, residual subset. The selection of thresholding techniques that differ in the criterion that it uses to decompose the dataset into main features and residual subsets. Most popular methods of thresholding are Heuristic rule, Kaiser rule and manual selection. Heuristic rule categorizes the principal components that collectively represent 95% of the variance of all principal components into the feature subset and the remaining into the residual subset. Kaiser rule categorizes the principal components that have higher variance than the average variance of all the principal components into the feature subset and the remaining into the residual subset.

Upon decomposition of the dataset into features and residuals subsets, the information that are embedded in the principal components must be expressed in terms of statistical indices to proceed with fault detection and diagnosis methods. Indices popularized by Qin consist of Hoteling's $\mathcal{T}^2$, Squared Prediction Error (SPE) $\mathcal{Q}$, and combined index φ. These indices provide a measure of how different a given measurement is compared to the defined baseline measurement. Therefore, defining the baseline measurement, which can be described as a healthy measurement, is a very important component in executing a FDD strategy using PCA.

Hoteling's index is calculated using Equation 2.23, where x_{new} is the new measurements, and $\widehat{D}$ is $\mathcal{T}^2$ projecting matrix. $\widehat{D}$ is defined in Equation 2.24, where $\widehat{\Lambda}$ is defined in Equations 2.25 and n is the size of the measurements. Hoteling's index calculates the distance between the feature subset of the baseline and the new measurement in the baseline's Principal component feature space.

$$\mathcal{T}^2 = x_{new}\hat{P}\widehat{\Lambda}^{-1}\hat{P}^t x_{new}^t$$

$$= x_{new}\widehat{D}x_{new}^t \tag{2.23}$$

$$\widehat{D} = \hat{P}\widehat{\Lambda}^{-1}\hat{P}^t \tag{2.24}$$

$$\widehat{\Lambda} = \frac{1}{n-1}\hat{T}\hat{T}^t \tag{2.25}$$

SPE index is calculated using Equation 2.26, where $\tilde{C}$ is the residual subspace that the new measurement is projected on. $\tilde{C}$ is defined in Equation 2.27. SPE index calculates

the distance between the new measurement and the residuals subset of the baseline measurement in the residual subspace.

$$SPE = Q = \|\tilde{x}_{new}\|^2 = \left\|\tilde{x}_{new}\tilde{C}\right\|^2 \tag{2.26}$$

$$\tilde{C} = \tilde{P}\tilde{P}^t \tag{2.27}$$

The combined index introduced by Yue and Qin is obtained using Equation 2.28 and 2.29, where δ is the upper limit of Q, τ is the upper limit of $\mathcal{T}^2$, and Φ is the projecting matrix. Equations 2.30 defines Φ. As shown in Equation 2.28, the combined index is a uniform combination of the previous indices.

$$\varphi = \frac{Q}{\delta^2} + \frac{\mathcal{T}^2}{\tau^2} \tag{2.28}$$

$$\varphi = x_{new}\Phi x_{new}^t \tag{2.29}$$

$$\Phi = \frac{\tilde{P}\tilde{P}^t}{\delta^2} + \frac{\hat{P}\hat{\lambda}^{-1}\hat{P}^t}{\delta} \tag{2.30}$$

The calculated indices are compared to an upper limit where exiting the limit indicates the presence of a fault condition in the system. Upper control limit of index $\mathcal{T}^2$, which is proportional to the F-distribution, is calculated using Equation 2.31 or 2.32 [53], [18]. $\propto$ is the assigned confidence level, n is the sample size of the original data, and l is the number of principal components. Equation 2.31 is utilized for larger sample size applications whereas, Equation 2.32 is more suitable for smaller sample size applications.

$$\tau^2 = \tau_\alpha^2 = \frac{l(n-l)}{n-l}\mathrm{F}_\alpha(l, n-l), where\ \forall \mathcal{T}^2,\ \mathcal{T}^2 \leq \tau^2 \tag{2.31}$$

$$\tau^2 = \tau_\alpha^2 = \frac{l(n^2-l)}{n(n-l)}\mathbb{F}_\alpha(l, n-l), where\ \forall \mathcal{T}^2,\ \mathcal{T}^2 \leq \tau^2 \tag{2.32}$$

For index $\mathcal{Q}$, which is quadratic in nature, the upper control limit is calculated using the weighted chi-squared distribution as shown in Equation 2.33, where $\mathcal{X}_\alpha^2$ is the chi-square distribution, m is the sample mean, v is the sample variance, g is the weight of the chi-squared distribution, and h is the degrees of freedom of the chi-squared distribution [54]. Equations 2.34 and 2.35 provide definitions of g and h.

$$\delta = \delta_{\mathcal{Q},\alpha} = g\mathcal{X}_\alpha^2(h), where\ \forall \mathcal{Q}, \mathcal{Q} \leq \delta^2 \tag{2.33}$$

$$g = \frac{v}{2m} \tag{2.34}$$

$$h = \frac{2m^2}{v} \tag{2.35}$$

Similar to index $\mathcal{Q}$, combined index, φ has been shown to be approximated using the chi-square distribution by Yue and Qin [15 ismail]. The upper control limit is described in Equation 2.36, where g_φ is the weight of distribution, h_φ is the degree of freedom and $\propto$ is the assigned confidence level. Definitions of the weight of distribution and degree of freedom are shown in Equation 2.37 and 2.38, respectively, where, $\theta_1 = trace\left(\widetilde{\Lambda}\right)$, $\theta_2 = trace\left(\widetilde{\Lambda}^2\right)$, and $\widetilde{\Lambda} = \frac{1}{n-1}\tilde{T}\tilde{T}^T$.

$$\zeta^2 = \zeta_\alpha^2 = g_\varphi \mathcal{X}_\alpha^2\left(h_\varphi\right), where\ \forall \varphi, \varphi \leq \zeta^2, \tag{2.36}$$

$$g_\varphi = \left(\frac{l}{\tau_\alpha^4} + \frac{\theta_2}{\delta_\alpha^2}\right) \Big/ \left(\frac{l}{\tau_\alpha^2} + \frac{\theta_1}{\delta_\alpha^2}\right) \tag{2.37}$$

$$h_{\varphi} = \left(\frac{l}{\tau_{\alpha}^{2}} + \frac{\theta_1}{\delta_{\alpha}^{2}}\right)^{2} \Big/ \left(\frac{l}{\tau_{\alpha}^{4}} + \frac{\theta_2}{\delta_{\alpha}^{4}}\right) \tag{2.38}$$

2.3.7 Multi-Scale Principal Component Analysis

As discussed in previous sections, PCA and wavelet transforms are very powerful tools that can be applied to FDD applications. More specifically, PCA is specialized in its ability to analyze multiple sensor measurements and identifying the relationships between those measurements whereas wavelet transform is specialized in its ability to decompose a time series measurement into time and frequency components. However, FDD strategy for complex systems such as an internal combustion engine requires the specialities of both PCA and wavelet transform simultaneously. In other words, for a successful FDD strategy, the decomposition of every measurement in time-frequency components and the identification of the correlation between all of the measurements at each time-frequency component is vital.

Bakshi introduced a new analysis method called Multi-Scale Principal Components Analysis (MSCPA), which can be described as a hybrid between PCA and wavelet transform. MSPCA consists of decomposing every measurement signal using discrete wavelet transform then performing PCA on the wavelet coefficients [55]. The concept of MSPCA is presented in Figure 15 where the signal is shown to be decomposed using DWT and the resulting wavelet coefficients become inputs to PCA where the loading and scores matrices are calculated for each wavelet coefficient. Upon loading and scores calculation, upper limits of $\mathcal{T}^2$, $\mathcal{Q}$, and φ are calculated for fault detection and diagnosis that were

introduced in Section 2.3.6. The difference between PCA and MSPCA is Principal components and the upper limits are calculated for each wavelet bin.

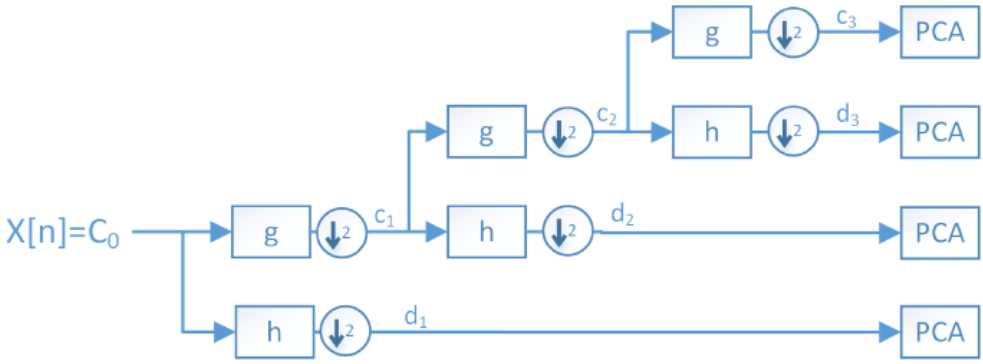


Figure 15: MSPCA schematic [12]

2.3.8 Mod-MSPCA

Yoon and MacGregor introduced a modified version of MSPCA called Mod-MSPCA [56]. The schematic of Mod-MSPCA is provided in Figure 16, which shows that the wavelet coefficients are not directly inputted into PCA upon decomposition through DWT. The wavelet coefficients in Mod-MSPCA are reconstructed using wavelet synthesis method then inputted into PCA at each level. The purpose of this modification is to utilize Reconstruction Based Charts for better detection and diagnosis of faults. RBC required sample size closer to the measurement sample size but due to the decomposition step prior to the PCA original MSPCA was not compatible with RBC. The reason for applying RBC for fault detection and diagnosis was to mitigate the issues of the upper limit technique such as inability to quickly detect changes in states of the operating conditions. Such limitation would manifest itself as false alarm indicating a presence of a fault even when

the fault has been resolved. Yoon and MacGregor states that RBC guarantees a higher accuracy in fault diagnosis. Further information regarding RBC can be found in Yoon and MacGregor's research [56].

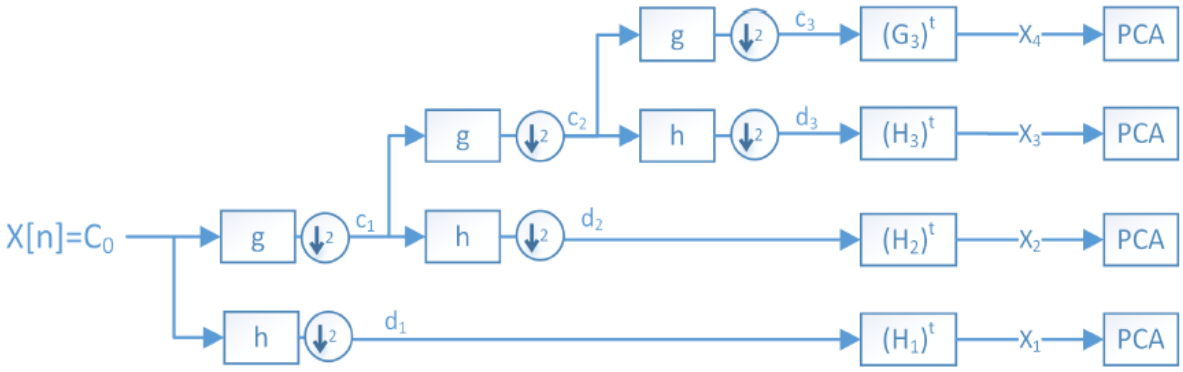


Figure 16: Mod-MSPCA schematic [12]

Chapter 3: Proposed FDD Strategy

This section will provide the details on the proposed fault detection and diagnosis strategy that will be developed and tested using real engine test-data. An overview of the purpose, constraints and assumptions involved in the development of the proposed FDD strategy is presented to provide a relevant summary of the information discussed in Chapter 2: FDD Strategies for Internal Combustion Engines as an introduction.

The purpose of the proposed FDD strategy is to monitor the internal combustion engine in a cycle-by-cycle basis to detect intermittent misfire faults that may occur in different cylinders of the engine. In this context, intermittent faults are faults that do not exhibit in every engine cycle. The application of the FDD strategy is aimed at detecting the presence of a misfire and the location of the fault condition, in other words, in which cylinder the misfire has occurred.

Due to the complexity of engine dynamics discussed in Section 2.2.1, the proposed FDD strategy utilizes the signal-based approach to enable rapid development of a robust strategy. The measurements used in the development of the FDD strategy includes pressure, vibration and angular position. The data acquisition for the development of the FDD strategy is envisioned to be incorporated smoothly into the existing engine development testing procedures. However, it is very important to note that the in-vehicle

application of the proposed FDD strategy is intended to only consider the vibration and angular position measurement. This is due to the very expensive cost associated with pressure sensors and the retrofitting of the pressure sensors in the engine as discussed in Section 2.2.2. The ability to detect fault conditions using only the vibration signal is hypothesized with an important assumption. This assumption is that there exists a correlation between the pressure and vibration measurement in the crank angle domain. With the knowledge of the pressure measurement's ability to assess the combustion quality, this assumption hypothesizes that the characterization of faulty combustion conditions is plausible with just the vibration measurement if the knowledge of the combustion quality obtained through the analysis of the pressure measurement is transferred through data labeling. Another important concept that will be utilized in the proposed FDD strategy is the analysis of data in the crank angle domain to correlate signal behaviors to specific component or engine event as discussed in Section 2.3.2.

The development method of the proposed strategy is outlined in Figure 17. It outlines four major components of the development method: equipment set-up & data acquisition, data pre-processing, feature extraction and classifier development. This chapter will introduce detailed concept of each component of the strategy. Chapter 4 and 5 will provide the detailed implementation of the concepts discussed in this chapter on real data set.

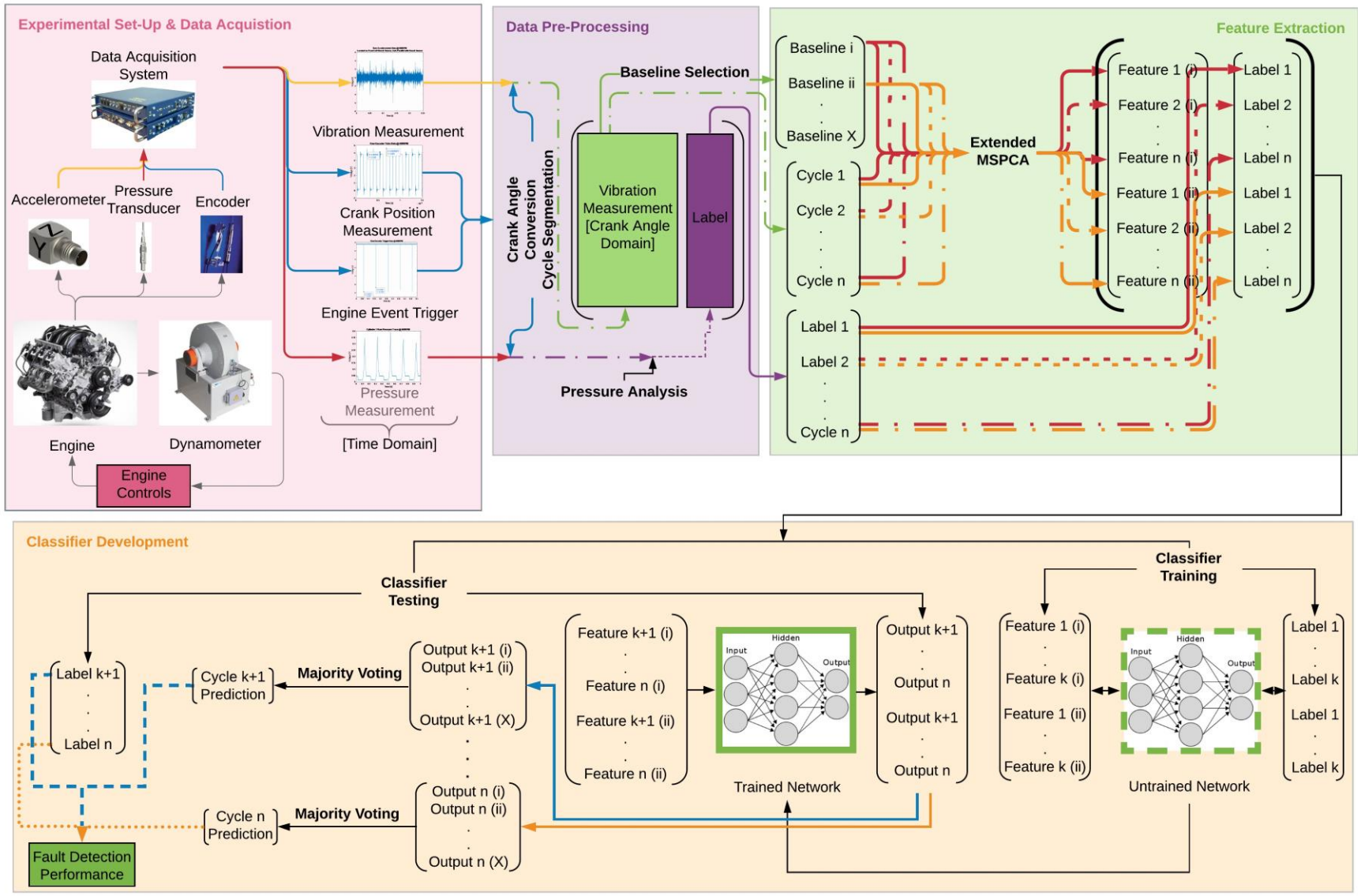

Experimental Set-Up & Data Acquistion
Data Acquisition System
Accelerometer
Pressure Transducer
Encoder
Engine
Dynamometer
Engine Controls
Vibration Measurement
Crank Position Measurement
Engine Event Trigger
Pressure Measurement
[Time Domain]
Data Pre-Processing
Crank Angle Conversion
Cycle Segmentation
Baseline Selection
Vibration Measurement [Crank Angle Domain]
Label
Pressure Analysis
Feature Extraction
Baseline i
Baseline ii
Baseline X
Cycle 1
Cycle 2
Cycle n
Extended MSPCA
Feature 1 (i)
Feature 2 (i)
Feature n (i)
Feature 1 (ii)
Feature 2 (ii)
Feature n (ii)
Label 1
Label 2
Label n
Classifier Development
Classifier Testing
Classifier Training
Label k+1
Label n
Cycle k+1 Prediction
Majority Voting
Output k+1 (i)
Output k+1 (ii)
Output k+1 (X)
Feature k+1 (i)
Feature n (i)
Feature k+1 (ii)
Feature n (ii)
Input
Hidden
Output
Trained Network
Output k+1
Output n
Feature 1 (i)
Feature k (i)
Feature 1 (ii)
Feature k (ii)
Untrained Network
Label k
Label 1
Cycle n Prediction
Output n (i)
Output n (ii)
Output n (X)
Fault Detection Performance

Figure 17: Development overview of the proposed strategy:

3.1 Proposed Experimental Set-up & Data Acquisition

As mentioned before, the proposed FDD strategy utilizes the signal-based approach due to the complex nature of the system being monitored. From the literature review, it can be concluded that modeling of engines requires tremendous amount of effort and resources that is not feasible for the development of a FDD strategy. Therefore, the use of engine testing process is suggested to develop a signal based FDD strategy for internal combustion engines. The benefit in utilizing the testing environment for the data acquisition component of the FDD development process is the amount of quality measurement that is available. These measurements have potential to provide valuable insight into the engine conditions and behaviors. The specifics of the engine and test cell set-up utilized in this research is provided in Section 4.1.1 and 4.1.2.

3.1.1 Proposed Instrumentation

The main measurement being proposed to be monitored in this FDD strategy is vibration in crank domain due to its low cost and effectiveness in characterizing faults in mechanical systems as discussed in Section 2.2.2. Another measurement of interest in the development of the proposed FDD strategy is the pressure traces of each cylinder. Pressure measurement is an example of measurement that is conveniently available in an engine testing environment. However, pressure measurement is not a measurement that is to be monitored in the application of the FDD strategy as pressure sensors are extremely costly and require complex retrofitting of the engine. Therefore, the pressure traces are proposed to be incorporated only in the development of the FDD strategy. The role of the pressure

traces during the development of the FDD strategy is to provide valuable information, specifically on the combustion quality of each engine cycle. This high quality information is to be incorporated into the development of the FDD strategy through date labeling method, where each vibration cycle is labeled as healthy or faulty through the analysis of the pressure measurement. The specifications and the locations of the accelerometers is further elaborated in Section 4.1.5. Similarly, the specifications of the pressure transducer are provided in Section 4.1.4. The pressure analysis method proposed is explained in Section 3.2 and the implementation of the method in this research is presented in Section 4.3.

The position measurement of the crankshaft of the engine is utilized to segment the continuous vibration and pressure measurement for the development process and just the vibration measurement for the application of the developed FDD strategy into aligned individual engine cycles. Also, the position measurement is to be utilized to convert the vibration and pressure measurement into crank angle domain from time domain. The rotational speed of an internal combustion engine is variable and never constant which means that events that are referenced in time provide little information regarding what process or component caused that event. However, when events are referenced using crank angle, simple analysis can infer useful information regarding the components or the process that caused the observed event. Detailed information regarding the optical encoder utilized is provided in Section 4.1.3. The theoretical concepts involved in the analysis of the encoder signals will be discussed in Section 3.2 and the implantation of the method is provided in Section 4.3.

3.1.2 Proposed Fault Induction Method

The purpose of data acquisition is to obtain engine behavioral data on healthy and faulty conditions under the same operating conditions to ensure that the fault condition is not being induced due to the change in the operating condition. The proposed method of fault induction during data acquisition process is to test the engine at parametric conditions that increases the possibility of a misfire. This method is proposed to obtain data that closely resembles the behavior that would be observed from an engine under normal operating conditions. Previously in Dohgri's research, misfires were induced by physically unplugging the spark plug from the engine [22]. This provides conditions of the engine that does not provide a realistic representation of how a misfire may occur in an engine that is operating within a vehicle. The details on the test equipment such as the engine, dynamometer, and the data acquisition hardware are provided in Section 4.1 and the testing conditions and test procedures are provided in Section 4.2.

3.2 Proposed Data Pre-Processing Methods

The purpose of the proposed data pre-processing method is to transform the raw signal that has been acquired through testing into a format that will allow the feature extraction method to extract the most information it can infer from the raw signals. This transformation includes the conversions of raw time series measurement into crank angle domain measurement and the segmentation of continuous measurement into individual engine cycles.

3.2.1 Proposed Crank Angle Conversion Method

As discussed in Section 2.3.2, representing the vibration and pressure measurements in crank angle domain is very important for reciprocating machines such as the internal combustion engine, where all events of the engine are intricately timed in terms of the crank position. The raw measurements that are obtained during the data acquisition is in the time domain and engine events are difficult to accurately be inferred in this domain due to wide operating speed ranges and its inability to hold a constant speed, which was explained in Section 2.1.1. The encoder measurement provides occurrences of a defined angular displacement of the crank shaft in time domain. The time measurements are proposed to be converted into crank angle domain using the crank angle and time relationship provided by encoder measurement through linear interpolation. Linear interpolation is suggested as the method of encoder measurement analysis as it is computationally very efficient compared to other interpolation methods. The detailed calculation of the proposed conversion method is provided in Section 4.3.

3.2.2 Proposed Cycle Segmentation Method

In order for the FDD strategy to monitor the engine in cycle-by-cycle manner, the continuous crank angle domain measurements need to be segmented into individual engine cycles. Engine cycle is defined as two revolution of the crankshaft which is equivalent of 720 degrees. Using the measurements in crank angle domain, 720 degrees of the crank shaft can easily be obtained. However, each individual cycle must be aligned to a specific engine event to ensure that alignment of the entire data set of cycles. Alignment is

especially important for feature extraction where time-frequency information is extracted. If the engine cycles of the same fault condition are not aligned at the same engine event, the time information that is extracted from the feature extraction process will differ between the features extracted from the two separate cycles. This means that the representation of the fault condition using both of these engine cycles can be considered invalid. For successful characterization of each fault condition, it is crucial that aligned engine cycles are utilized for fault extraction. The engine event that is utilized for this alignment is the top-dead-center event of cylinder 1. The detailed application of this proposed cycle segmentation method is provided in Section 4.3.

3.2.3 Proposed Pressure Analysis Method

The combustion quality information of the engine cycles inferred from the pressure traces using thermodynamic theories are to be incorporated in the development of the proposed FDD strategy. The method of pressure analysis that is proposed in this research is the calculation of Indicated Mean Effective Pressure (IMEP). The purpose of this analysis is to define a misfire mathematically using thermodynamic theories. IMEP represents the average pressure that is acting on the piston during the engine cycle. IMEP is a very accurate and an easy to calculate indicator of misfires. For the pressure trace of a combustion chamber when an engine is motoring, which means that there are no combustion occurring inside the combustion chambers, the IMEP value is zero. A calculated value of IMEP below a set threshold limit will be defined as a misfire Therefore, if an IMEP of a combustion chamber during normal operating conditions of an engine is a value close to the threshold limit, it would indicate that a misfire has occurred in that

combustion chamber. The steps involved in calculating the IMEP values of each engine cycle is outlined in Section 4.3.

3.2.4 Proposed Data Labeling Method

The information regarding the combustion quality obtained through the pressure analysis is to be incorporated into the development strategy through a process called labeling. Labeling of vibration data for each engine cycle is required as the fault conditions, which are differentiated by its combustion quality, are not easily inferred using vibration signals. With the assumption that correlation exists between the vibration measurements and the corresponding fault condition, the label is aimed at guiding the characterization of each condition using vibration measurements. The true state of the combustion quality is to be determined through the pressure analysis discussed in the previous section. Different fault labels were given to misfires occurring in different cylinders to allow the classifier to be able to differentiate a misfire that has occurred in one cylinder or another, for fault diagnosis component of the FDD strategy. The main purpose of the labeling process is to eliminate the need for using pressure transducers in vehicles for the application of the FDD strategy. The process of the labeling method is provided in Section 4.3.

3.3 Proposed Feature Extraction Method

Feature extraction of vibration measurements of each segmented and aligned engine cycle that are labeled using pressure measurements are proposed to be performed using Extended Multi-Scale Principal Component Analysis (Extended-MSPCA).

Extended-MSPCA, which was introduced by Ismail, is a feature extraction method that utilizes PCA and wavelet packet transform (WPT). This method is used due to its ability to analyze non-transient behaviors of a complex system that have been captured through multiple sensors. In short, the Extended-MSPCA decomposes a given signal using WPT, which provides corresponding wavelet coefficients then applies PCA on those coefficients.

Extended-MSPCA is very similar to Mod-MSPCA, which was discussed in Section 2.3.7. However, the feature that differentiates the Extended-MSPCA from Mod-MSPCA is improved performance in the characterization of the fault conditions. The main difference between the two methods is that Extended-MSPCA utilizes Wavelet Packet Transform (WPT) while Mod-MSPCA utilizes Discrete Wavelet Transform (DWT). The mathematical theories regarding WPT and DWT were discussed in Section 2.3.5.2. The two transforms differ in the way they decompose the signal being transformed. DWT decomposes the signals in a way that highlights the higher frequency contents whereas, WPT decomposes the signal equally along the signal's frequency spectrum. Therefore, WPT provides better granularity in terms of frequency in comparison to the DWT. The Extended-MSPCA is proposed, as opposed to the Mod-MSPCA, due to its ability to highlight all of the frequency bands equally. As faults may occur in both high and low frequencies, it is very important to utilize the feature extraction method that is capable of capturing both those cases.

In addition to the type of wavelet transform utilized, Extended-MSPCA differs from Mod-MSPCA in the way normalization is achieved. In Mod-MSPCA the raw signal is immediately decomposed using DWT and the output coefficients are normalized for PCA. However, Extended-MSPCA normalizes the raw measurements before decomposing them using WPT. Therefore, the outputs of WPT, which are the wavelet coefficient, are considered to be normalized which means that normalization upon WPT is unnecessary for PCA. The summary of the proposed normalization method is provided in Figure 18 and Figure 19. The proposed order of normalization allows for features extraction process to minimize the effects of varying scale and variances of measurements due to difference in the sensor's calibration or sensitivity.

Figure 18: Mod-MSPCA signal transformations [12]

Figure 19: Extended MSPCA signal transformations [12]

Given the difference between Mod-MSPCA and Extended-MSPCA, the detailed steps involved in executing the Extended-MSPCA is as follows. Upon data acquisition and necessary data pre-processing that were introduced previously in this chapter, the baseline measurement is to be first defined. The Extended-MSPCA generates the features for a given measurement by comparing how different the new measurement is in comparison to a defined baseline measurement. This baseline measurement is proposed to represent a

normal or healthy measurement. However, due to the variability that is inherent in engine operation, which was discussed in Section 2.1.1, the selection of the baseline measurement is not trivial. It is clear that the attempt to represent what normal engine cycle looks like with a single-engine cycle measurement is a contradiction in itself. If a single-engine cycle is to be plotted in the feature domain, the normal cycle will be represented as a single point. If all of the normal cycles were to be plotted in the feature domain, the normal cycle will be represented as a cluster of points. Given that cycle-to-cycle variations exists even in normal conditions; the cluster representation is a more realistic depiction of what normal conditions look like as opposed to the single point representation. Therefore, the baseline measurement is proposed to be represented by multiple healthy measurements as opposed to a single healthy measurement. The process of baseline selection is described in Section 5.1. Multiple baseline measurements are not proposed to be manipulated to provide a calculated baseline data such as an average of the multiple baseline measurement. Instead, the utilization of the multiple baseline measurements is proposed to extract multiple representations of the measurement being observed. For example, if 10 individual cycles are selected to represent the baseline measurements of the given system.

Upon selection of the baseline measurements, the measurement, that the feature will be extraction from, is proposed to be normalized using Equations 3.1 and 3.2 where normalization is performed based on the parameters of the assigned baseline measurement. The assigned baseline measurement is also normalized using its own parameters.

$$signal_{normalized} = \frac{signal_{raw} - mean\ value\ (baseline\ center)}{variance\ (baseline\ raw\ signal)} \tag{3.1}$$

$$x_{normalized} = \frac{x_{raw} - \frac{1}{n}\sum_1^n x_{raw,i}}{\frac{1}{n-1}\sum_1^n \left(x_{raw,i} - \frac{1}{n}\sum_1^n x_{raw,i}\right)^2} \tag{3.2}$$

Once the measurements are normalized, it is decomposed into time-frequency domain using WPT. The PCA component of the proposed feature extraction method is then applied to calculate the principal component matrices for the healthy and the faulty measurements using Equations 3.3 and 3.4 given the relationship provided in Equation 3.5.

$$\Sigma_{T,healthy} = COV(T_{healthy}) \tag{3.3}$$

$$\Sigma_{T,faulty} = COV(T_{faulty}) \tag{3.4}$$

$$COV(T) = E(TT^t) - E(T)E(T)^t \tag{3.5}$$

A statistical index was introduced by Ismail to enable fault diagnosis capabilities of the Extended-MSPCA [12]. This index, which is defined by Equation 3.6, provides a quantitative measure of a fault condition.

$$F_{c,j} = sign\left(L_j\right) \circ \sqrt{\sum_{i=1}^{m} sign\big(COV(F)\big)_j \circ [COV(F) \circ COV(F)]_j} \tag{3.6}$$

Further to the above detailed overview of the theory related to Extended-MSPCA, specific application method is proposed. This application method is heavily focused around

how to perform feature extraction using multiple baselines. The proposed feature extraction process for the development of the FDD strategy aims to produce multiple representations of the fault signature based off of multiple baseline data as shown in Figure 17. This will lead to a clustering of healthy and faulty data in the feature domain. More representation of one faulty data with respect to multiple baseline data allows for the classification algorithm to account for the variable nature of the engine cycle. It should be noted that these baseline data are not used to represent a condition. In other words, a feature representation of baseline data is not extracted or used in the training process. It is also important to have healthy data represented in comparison to the multiple baseline data to account for the variability even within the healthy condition.

3.4 Classifier Development

In this FDD strategy, a supervised machine learning algorithm is proposed to develop a classifier algorithm that would classify features into fault categories. Supervised learning algorithms take labelled data set that provide input-output examples and learns the relationship between the input features and the output labels [57]. During the learning process, the algorithm refines the mapping of the inputs and the outputs by minimizing the error between the desired outputs and the classifier algorithm's output. The errors are represented by a cost function and it is minimized by updating the weights.

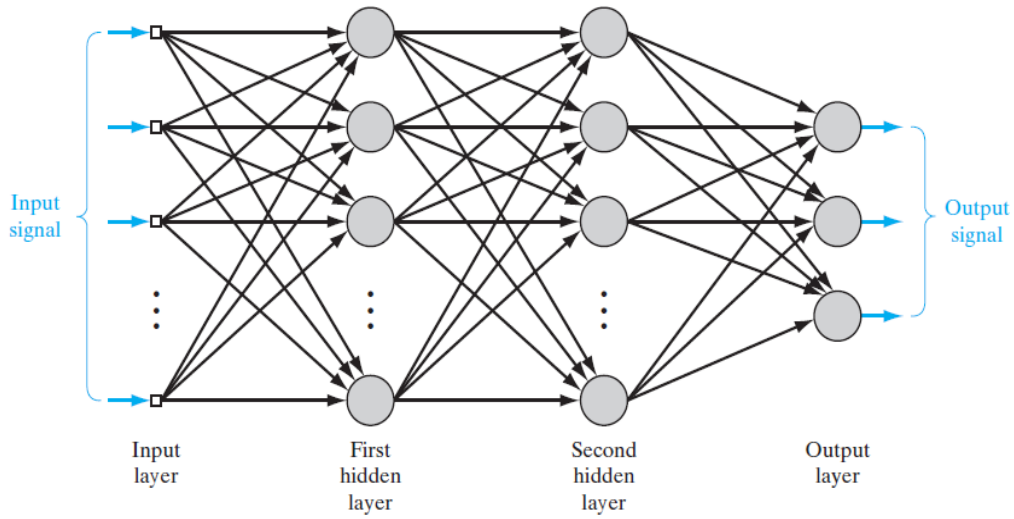


Figure 20: MLP with two hidden layers [57]

A type of artificial neural network (ANN) called a multilayer perceptron (MLP) is proposed to be the classifier algorithm for this research due to its compatibility with nonlinearity and ease of application. Similar to the human brain, ANN are composed of fundamental information processing units called neurons. MLP consists of multiple clustering or a row of neurons called a layer, as shown in Figure 20. The simplest MLP consists of three layers: input, hidden and output layer. Figure 20 depicts MLP architecture with two hidden layers. It is important to note that every neuron from a layer is connected to every neuron in the previous and the following layer, this type of architecture is known as a fully connected neural network. Figure 20 also shows that the information flows from the input to the hidden layers then to the output layer, this type of architecture is known as a feedforward neural network.

Model of a neuron, which consist of m number of inputs, a bias, an activation function and an output, is depicted in Figure 21. As shown in the figure, each neuron receives m number of inputs which are multiplied with their respective weights. The weights correspond to the connection between a neuron from the previous layer and a neuron in the following layer. The weights are unique to each connection and are updated during the training process. Figure 21 also shows that each neuron consists of its own bias which also gets multiplied with its corresponding weight value. The values obtained through the weight multiplications are summed at the junction as $v_k = \sum_{j=0}^{m} w_{kj} x_j$, which is fed into the activation function as an input to calculate the output of the neuron, $y_k = \varphi(v_k)$ [57]. A variety of activation functions are available such as sigmoid, tangent hyperbolic, softmax and rectified linear units. The nonlinear behavior of these activation functions enables artificial neural networks to be nonlinear. As this research involves a multi-class classification problem, a softmax activation function is used in the output layer.

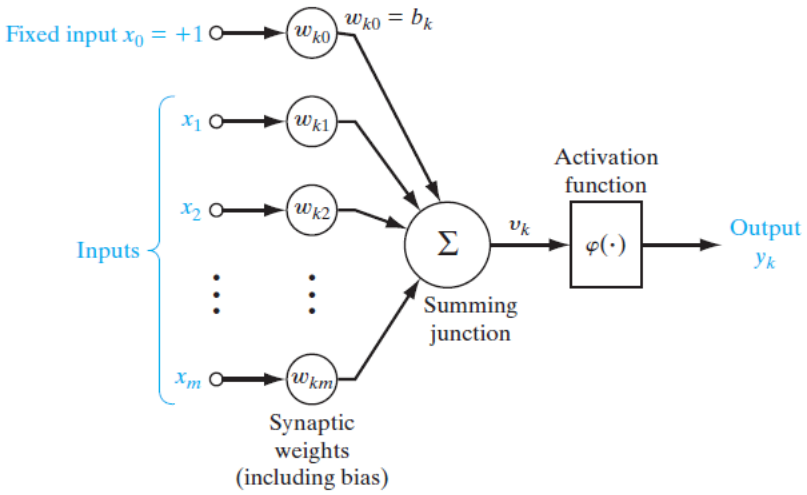


Figure 21: Model of a neuron [57]

A MLP is trained using a technique called backpropagation. During training there are two types of signals that flow through the network which are called function and error signals as shown in Figure 22 [57]. The main distinction between the two signals are the direction in which they travel through the network. The forward propagation is carried out by the function signal and the back propagation is carried out by the error signal [57]. During forward propagation, the function signal travels through each layer from one neuron to another to transform the input into an output value. During back propagation, an error signal is obtained by comparing the output value of the neural network to the desired output. This error signal travels backwards through the network layer by layer. As the error signal travels, the weights associated with each neuron to neuron connection is adjusted to minimize the error [57]. As mentioned before, errors are represented by loss functions which are minimized by a gradient method. The feedforward and backpropagation is repeated until the loss function is minimized to a satisfactory value. For this research a cross-entropy loss function is proposed to be minimized using the scaled conjugate gradient backpropagation methods.

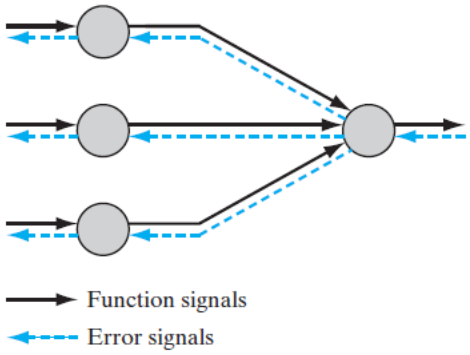


Figure 22: Function and error signals [57]

As mentioned before, the multiple baseline strategy is proposed in this research to take into consideration the variable engine dynamics. Given a single engine cycle, n individual features are extracted if n number of baselines are used. These individual features are different feature representations of a single engine cycle. During the training of the classifier, it is suggested that each representation of a given fault is to function as an individual training sample. The classifier algorithm will not know that these representations were obtained from one engine cycle. This method is proposed to provide classifier with enough variations of each fault condition to learn from. Exposing the classifier to as many variations possible allows the classifier to understand the variable nature of the engine behavior. This is especially useful if the dataset obtained is small.

Once the classifier is fully trained, the FDD strategy will utilize a majority voting method to detect a fault. For example, if n number of baselines are used to extract features from a new engine cycle, n number of feature representations will be extracted. Each of the feature representations will separately be classified using the trained classifier, which will

provide n different assessments of the fault condition. A majority voting method is utilized to make the final prediction on the fault condition of the new engine cycle. The concept of the majority voting method is to select, among all the of the different predictions that were provided by individual features, the prediction that had the most votes. The schematic process of this method is shown in Figure 17 under the classifier development subset. This majority voting method is utilized to reduce detection error caused by uncertainty of the fault condition dimensions in the feature space.

Chapter 4: Data Acquisition

In this section, the details of the proposed strategies regarding data acquisition are provided. These subsections include the experimental setup, the engine test procedures and data pre-processing.

4.1 Experimental Setup

This chapter provides detailed descriptions of the experimental setup utilized for the development of the FFD algorithm for internal combustion engines. The FDD algorithm monitors the engine on a cycle by cycle basis. It is a signal-based approach that detects intermittent combustion related to faults in different cylinders.

The experimental setup was used to acquire data for feature extraction of different conditions, training of the machine learning classification tool, and testing of the classifier. The main elements of the experimental setup include a test cell, an engine, an engine dynamometer, a data acquisition system, and sensors (including an encoder, pressure transducers and accelerometers). The data acquired includes measurements from the encoder, accelerometers, and pressure sensors in time domain.

4.1.1 Engine

All the experiments for data acquisition were performed on a Ford's V8 spark ignition, pushrod developmental engine. The engine was provided by Ford's Powertrain Engineering Research and Development Center (PERDC). Detailed specifications of the engine are omitted due to confidentiality. However, the focus of this research and the resulting conclusions are not dependent on the specifics of this particular engine. Instead, the measurements acquired and the strategy implemented consider phenomena that are universal to engines of this type, at the least; V-style, 8 cylinder, and overhead valve model. Figure 23 illustrates the difference between an overhead valve and an overhead cam engine design and Figure 24 illustrates the difference between an inline engine and a V-style engine.

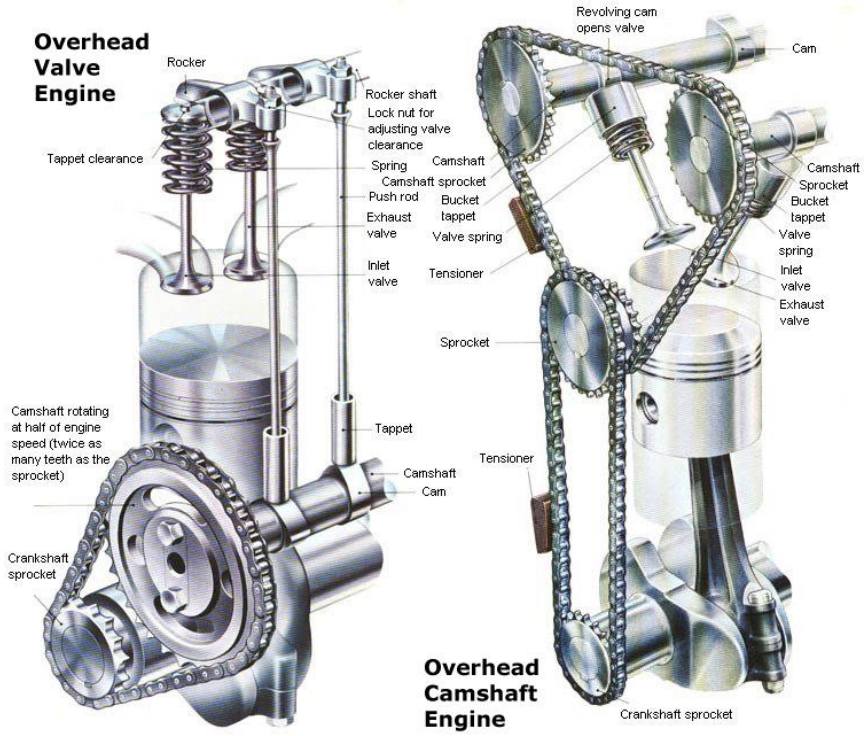


Figure 23: Overhead valve vs overhead cam engine design [58]

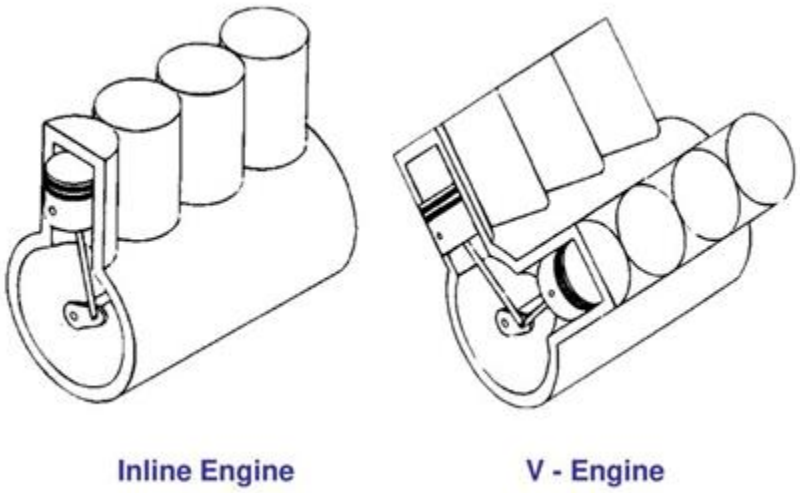


Figure 24: Different styles of engines [59]

Figure 25 shows the cylinder numbering for the test engine utilized in this research.

The firing order of the engine is cylinders 1-5-4-8-6-3-7-2 as shown in Figure 22.

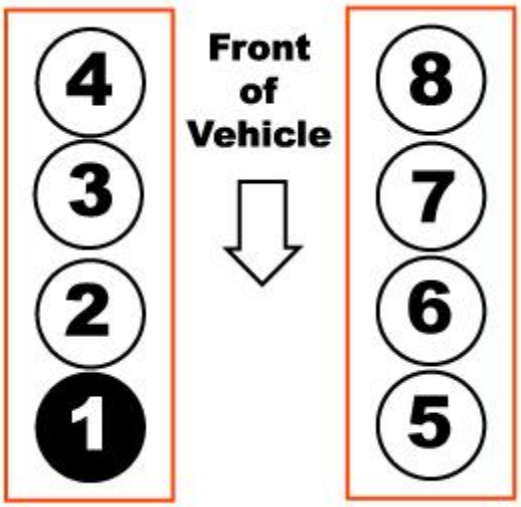


Figure 25: Engine cylinder numbering for the V8 test engine [60]

4.1.2 Test Cell

The testing facility was provided by Ford Powertrain Engineering Research and Development Center. The test cell is composed of an engine dynamometer, engine assembly and monitoring instrumentations. The test cell that was utilized for this research was unique amongst all of the test cells that are located at PERDC. The test cell is in a semi-anechoic chamber which ensures the elimination of surrounding noise and absorption of the energy created within the chamber. In addition, only the engine assembly, including the engine, was housed in the test cell to ensure that vibrations from the engine alone are obtained during testing.

4.1.2.1 Engine Assembly

The engine assembly consisted of the clutch, transmission assembly, fan belt, power steering pump, alternator, compressor, powertrain mounts, and mount brackets. The purpose of the engine assembly is to simulate the in-vehicle conditions that the engine is to operate within.

4.1.2.2 Dynamometer

The dynamometer was housed outside the chamber in a separate utility room. This separation ensures that only the noise and vibration from the engine assembly is being captured during data acquisition without external noise. The following table summarizes the specifications of the dynamometer that was utilized in this research.

Table 1: Dynamometer Specifications

Manufacturer	Meidensha Inc
Manufacturer Date	2001
Type	FREC Dynamometer
Model	S95-355L
Number of Poles	2 P
Base Speed	2500/2647 RPM
Max Speed	8000 RPM
Absorbing Power (Continuous)	370 kW
Motoring Power (Continuous)	250 kW

4.1.3 Encoder

The encoder is used to measure the rotational position of a shaft [61]. An encoder is equipped with an optical sensor, light source and rotating disk with either opaque and transparent sections or slits. The optical encoder is mounted on the rotating shaft to be studied and the optical sensor outputs “on” or “off” pulses. depending on whether or not it sensed the light that was emitted by the light source. The optical sensor would sense the light in the case where the transparent or the opening of the slit is passing through the light source and the sensor which are in line with each other with the disk rotating in between the two as shown in Figure 26. The output will be “off” if the light is blocked due to the opaque section or the solid portion of the disk. The output of the encoder is a digital signal which means that it will be in the form of a square wave with the high values indicating the “on” state and the low values indicating the “off” state. With the knowledge of the angular width of the sections or the angular distances between slits, the angular position of

the shaft can be obtained with respect to time. The angular speed and acceleration of the shaft can also be derived from the position and time data obtained from the encoder.

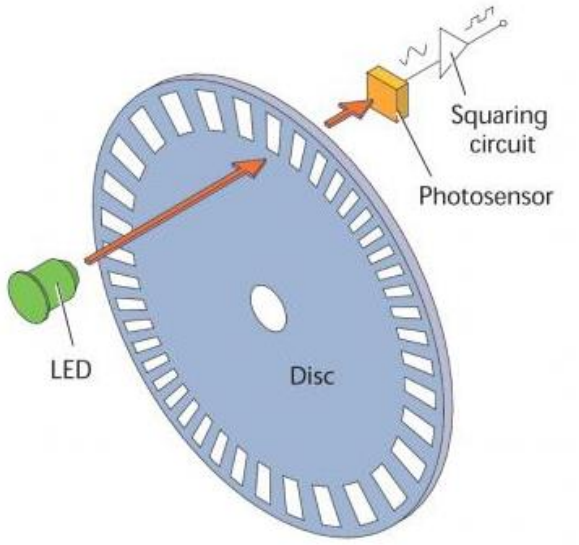


Figure 26: Encoder components [62]

In this experiment, the AVL 365 angle encoder was used as it provides high precision and high resolution angle information that is required for converting the transducer signals from time domain to crank angle domain. This encoder has a resolution of 0.5-degree crank angle which means that each engine cycle, which involves two rotations of the crankshaft, will consist of 1440 data points. A trigger signal was also obtained from this encoder which indicates the occurrence of the top dead centre (TDC) of Piston 1 with an offset. The offset of the trigger signal to TDC of Piston 1 is 131.662 degree which means that the trigger signal gets switched "on" 131.662 degrees after the occurrence of TDC of Piston 1. The trigger signal is crucial in resampling of the transducer signals as it ensures that every cycle is synchronized properly in crank angle domain and

allows for the data set to be divided into cycles in the time domain. The AVL 365 angle encoder has the RS422 transistor-to-logic (TTL) output driver, output range of 0 – 5V and with a standard threshold being 3.5V for "on" signal [63].

4.1.4 Pressure Transducer

Pressure traces provide useful insight into the combustion happening inside the engine. To correctly label the data, pressure traces were used to calculate values such as Indicated Mean Effective Pressure (IMEP) to determine if a combustive fault has occurred or not in a given engine cycle.

Two main types of transducers for sensing pressure are piezoelectric or piezo-resistive. The application determines which type of sensor should be used. These two types of transducers differ in that piezoelectric pressure sensors measure the electrical charge produced by the piezoelectric material, whereas piezo-resistive pressure sensors measure the change in electrical resistance of a silicon semiconductor.

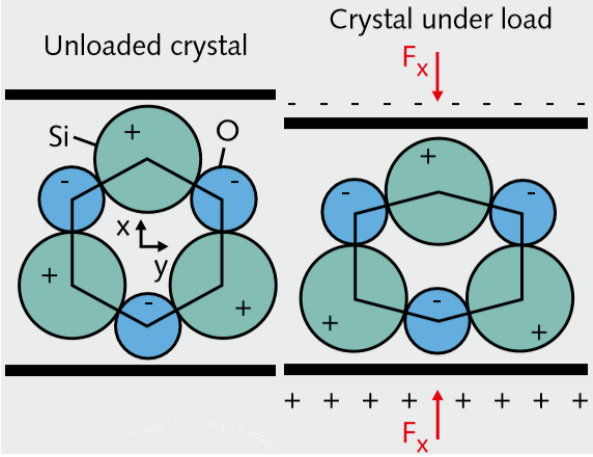


Figure 27: The mechanical load on the crystal producing electrical charge through electric dipole [64]

For the piezoelectric sensor, the electrical charge is produced by the piezoelectric material, such as quartz, and occurs when the surfaces of the crystal are loaded mechanically. This load causes the shift of positive and negative crystal lattice elements in relation to each other resulting in the formation of an electric dipole [64]. These sensors have a very high natural frequency which makes them ideal for measuring fast pressure rise times. However, one of the downsides of piezoelectric sensors is that the measurement drifts upon application of a static load.

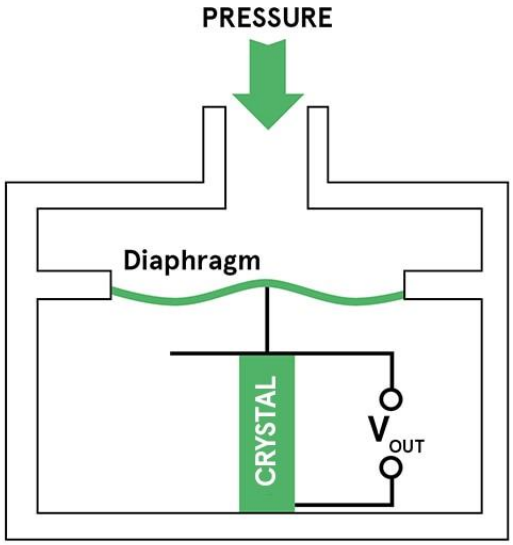


Figure 28: Piezoelectric pressure sensor components [65]

For the data acquisition portion of the research, pressure traces were obtained through Kistler's PiezoStar Type 6125C pressure sensors which are designed for in-cylinder pressure measurement of combustion engines. Ford PERDC provided cylinder heads with machined holes for each cylinder that are compatible with the pressure transducers for testing. PiezoStar pressure transducers are piezoelectric sensors with PiezoStar crystals which are produced by Kistler instead of quartz. It has a measuring range

of 0 – 300 bar and a sensitivity of -33 pC/bar [66]. The negative sensitivity of the pressure transducer is due to the fact that pressure applied to a piezoelectric sensor produces negative charge signal. This phenomenon is not found in the output signal as external charge amplifier converts the negative going charge into a positive voltage signal [64]. Piezoelectric transducers were utilized instead of the piezo-resistive transducers as the nature of combustive pressure signals are highly dynamic in nature.

4.1.5 Accelerometers

. Accelerometers are used to capture vibrational behaviors by measuring the dynamic acceleration of a mechanical component or assembly, [67]. Similar to pressure transducers that were discussed in the previous section, accelerometers use piezoelectric material to measure acceleration. An accelerometer consists of a seismic mass, piezoelectric material and preload bolt. The acceleration experienced by the component being studied is transmitted to the accelerometer and causes the seismic mass inside it to also accelerate. A proportional force generated by the acceleration of the seismic mass is induced on the piezoelectric material that is clamped between the seismic mass and the base of the accelerometer by a preloaded bolt. The proportional force applied to the piezoelectric material by the seismic mass results in an electric dipole formation due to the shift in the lattice structure of the piezoelectric material. This electric dipole then generates electrical charge proportional to the acceleration that can be acquired by the data acquisition hardware.

There are two types of accelerometers, charge type piezoelectric sensors and integrated electronic piezoelectric (IEPE) sensor and they differ in the way they convert the charge from the piezoelectric crystal. In the charge type accelerometers, the charge is converted through an external amplifier or inline charge converter. However, an IEPE sensor does not require an external amplifier as it has a built-in charge amplifier. The IEPE sensors require a data acquisition hardware that provides current excitation for the built-in amplifier.

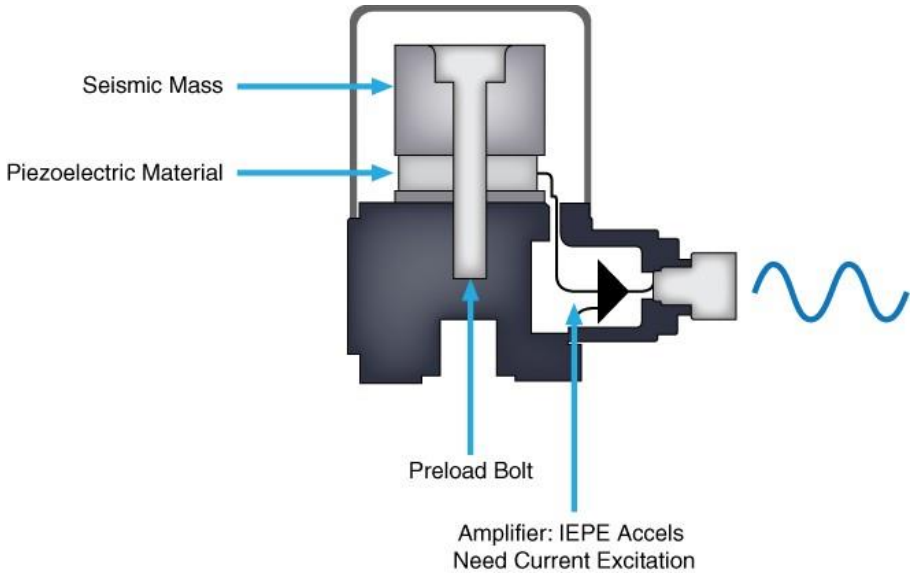


Figure 29: IEPE accelerometer components [67]

In this research, Dytran IEPE tri-axial accelerometers were used to obtain the vibration from various parts of the engine. The accelerometer has a range of +/- 5000g, frequency response of 10 kHz and a sensitivity of 1mV/g [68]. It should be noted that some may argue that charge type accelerometer would have been preferred over the IEPE as they are suited for high temperature applications. However, given the fact that the normal operating temperature for an engine is anywhere between 195 to 220 degrees Fahrenheit

and that the operating temperature of the IEPE accelerometers are rated up to 250 degrees Fahrenheit, IEPE sensors are used with cautionary measures in place such as wrapping the exhaust and sensor cables with heat shield wraps.

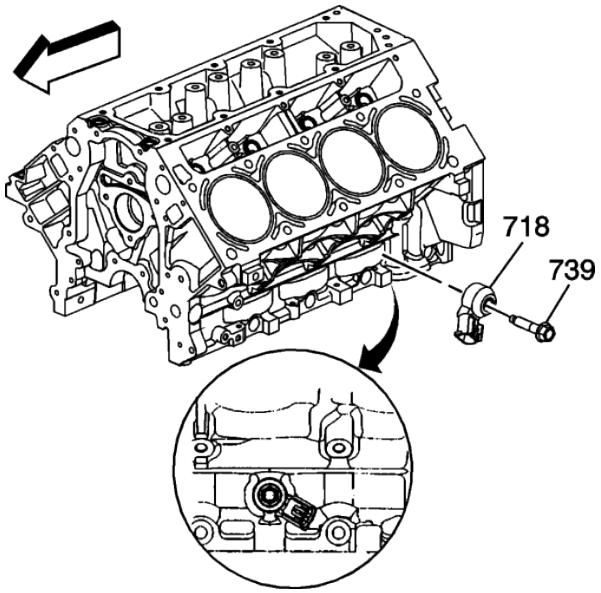


Figure 30: Knock sensor locations on sample V8 engine [69]

The accelerometers were mounted on the engine with epoxy or with bolts when possible to ensure that the true vibrational readings are being obtained. Two accelerometers were placed on top of two knock sensors in the left bank of the V8 engine. From the accelerometer placed on top of the knock sensors, only the z-axis measurement was acquired. The z-axis is parallel to the vibrational measurement obtained by the knock sensor. Therefore, this measurement should be indicative of the vibrational components that knock sensors observe. This location is strategically selected as the success of the proposed FDD strategy using sensors and sensor locations that already exist in the manufacturing process would reduce cost of implementing such strategy. The locations of

the knock sensors for the experimental engine are shown on a similar V8 engine in Figure 30. Two other accelerometers were placed in the valley of the V-style engine as shown in Figure 31. The placement of these sensors was solely explorative which aims to capture the overall vibrational behaviors of the engine. All 3 axes from these accelerometers were acquired to observe the up and down, left to right and front to back vibrations of the engine.

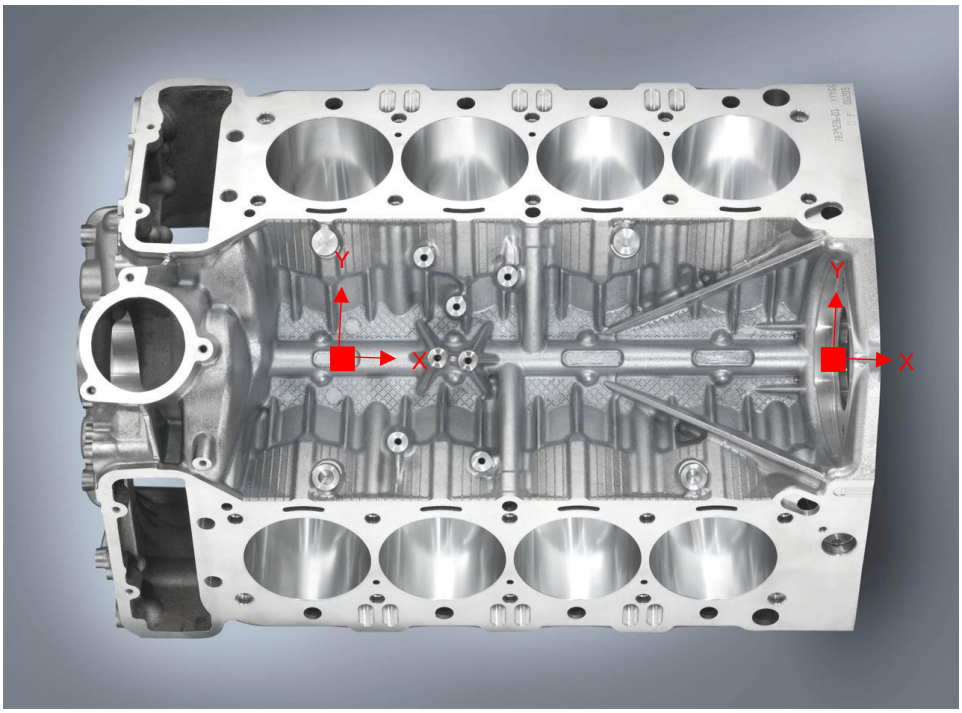


Figure 31: Accelerometer locations on sample V8 engine block [70]

All the accelerometers were calibrated to 100 Hz at 1g-rms as specified by Dytran prior to mounting. The calibrations were performed using The Modal Shop 9100D Portable Shaker Table.

4.1.6 Data Acquisition Hardware

During a data acquisition process, sensors convert physical phenomenon to electrical signals which are collected and digitized by a data acquisition hardware [71]. Depending on the data acquisition hardware, the module may be capable of providing either analog input/output, digital input/output or counter operation functionality on a single device or all of the listed functionalities on a single multifunction device. The type of sensor dictates the functionality of the data acquisition hardware required to acquire the signal [71]. For example, the encoder signal which outputs a digital signal would require the digital input/output functionality whereas, the accelerometer signal which outputs an analogue signal would require the analogue input functionality and Analog-to-Digital Conversion (ADC) to acquire the data.

For this research experimentation, signals from the encoder, accelerometers and pressure transducers were acquired by multifunction devices called PROSIG P8020 and P8012. These two modules allow for fixed channel counts of 24 low speed 24-bit ADC channels and 16 high speed 16-bit ADC channels to be recorded simultaneously. These multifunction devices have a range of +/- 10 V and an accuracy of +/- 0.1% of the range. The high speed and low speed modules are able to support a maximum sampling frequency of 400k and 100k samples/sec/channel, respectively. Given the maximum engine speed that the data will be acquired at, frequency component of the encoder signal was calculated.

Although Nyquist Theorem suggests that a signal can be reconstructed with a sampling rate equal to twice the highest frequency of the interested component, National

Instruments advise that a sampling rate equal to ten times the highest frequency of the interested component should be used [72]. The sampling rate of the encoder was carefully selected to ensure that the angular position of the crankshaft is being captured as precisely as possible without any loss of information. All the vibration and pressure signals were recorded at 100kHz/channel and the encoder data was recorded at 400kHz/channel. The encoder signal was sampled at a much higher rate compared to the transducers as the encoder's data will be used to resampled the transducer data in crank angle domain. It is very important to have data point that is as close to the timestamp as possible to when the switch occurred. Therefore, in this application a high sampling frequency of the acquisition system is preferred.

Prosig data acquisition system performs filtering on the signals that are to be obtain through a built-in anti-aliasing filter. The cut-off frequency for the filtering is adjusted automatically according to the sampling rate of the signals to ensure that aliasing effects are not captured in the signals acquired.

4.2 Engine Test Procedure

4.2.1 Operating Conditions

The fault of interest for this research is intermittent misfire. The associated FDD experiment involves obtaining both healthy and faulty cycles under a same operating condition and by running the engine in a condition where the chances of the fault occurring are high. Under the normal operating range of the engine, the idle condition is considered to be the most unstable, with significant fluctuations in combustion quality. Heywood

specifies that at idle conditions, chances of partial-burning cycles are higher, [4]. Therefore, the data was collected at the idle condition where the engine is running without a load at 600 rpm. It is also mentioned that the chances of misfires can be reduced at the idle condition by retarding the spark timing [4]. Therefore, spark timing was advanced incrementally to increase the severity of the engine's instability. The following list summarizes the operating condition:

- idle speed of 600 RPM
- no load
- engine oil temperature below 265 F
- coolant temperature below 195 F
- spark timing advance varied from 35 to 55 degree

4.2.2 Testing Procedures

The engine was warmed up at 1500 rpm with a load of 2 bar until the oil temperature reached 160 F, and a stable condition is reached. Once the engine is warmed up, the engine speed was lowered to 600 rpm, the load was removed and the appropriate spark timing advance was applied. The data acquisition was then triggered when steady state was reached further to changes made to spark timing advance.

4.2.3 Raw Data Collected

Accelerometer Data

Measurement signals from eight accelerometers were recorded in this research. Raw accelerometer data from a three axes accelerometer located on the front valley is

shown in Figure 32 to Figure 34. Raw accelerometer data from an accelerometer located on the front left knock sensor is show in Figure 35.

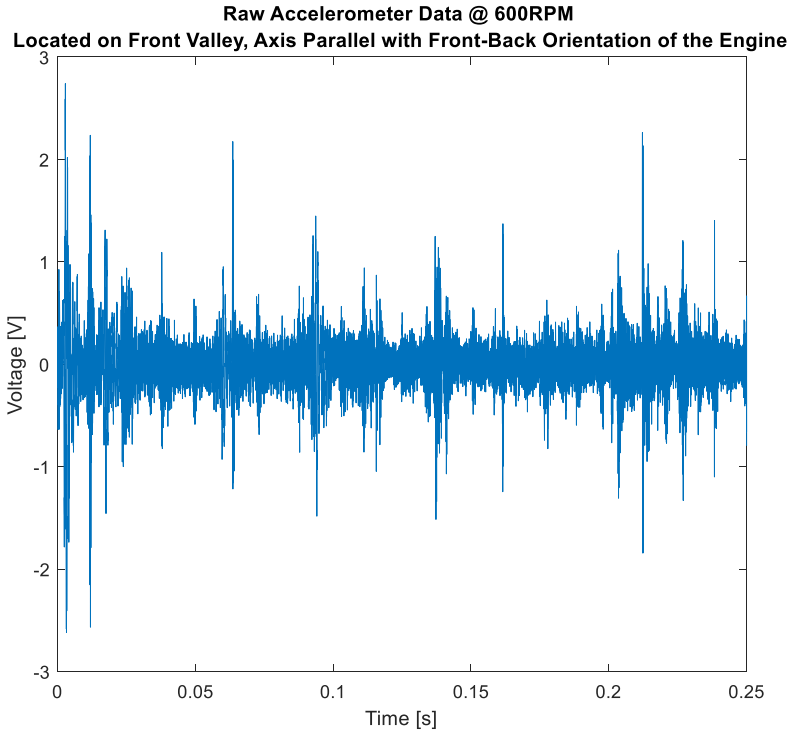


Figure 32: Raw accelerometer data from front valley, axis parallel with front-back orientation of the engine

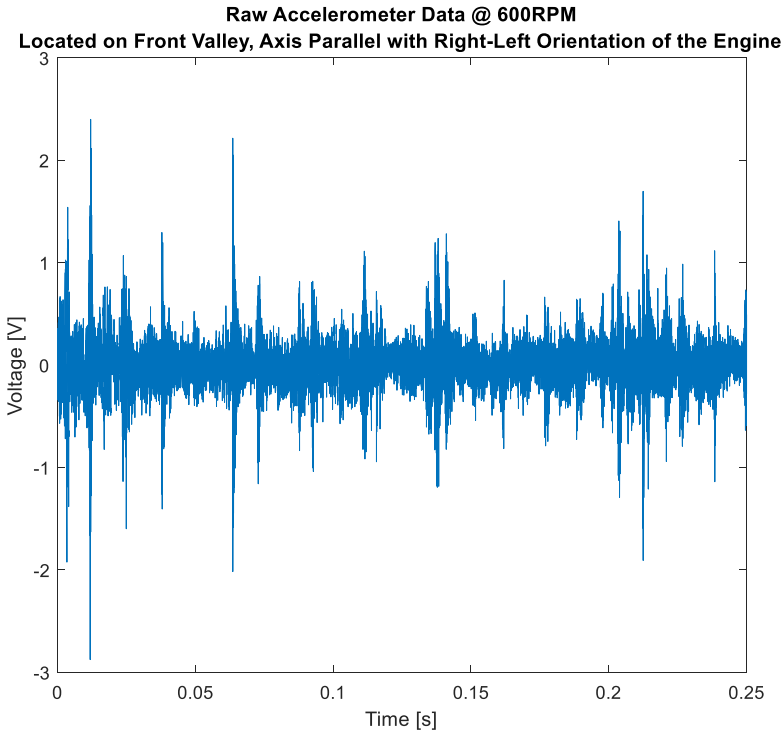


Figure 33: Raw accelerometer data from front valley, axis parallel with right-left orientation of the engine

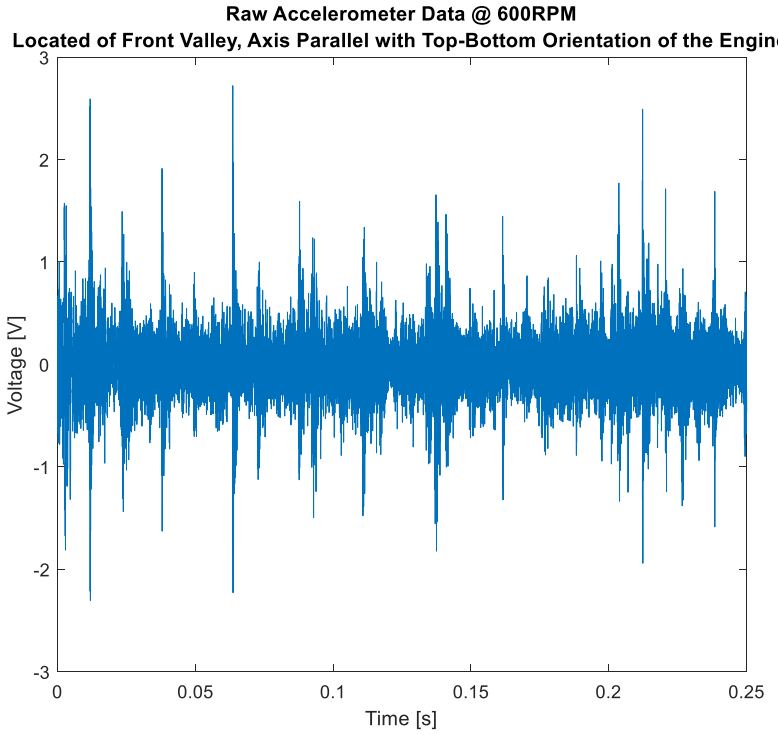


Figure 34: Raw accelerometer data from front valley

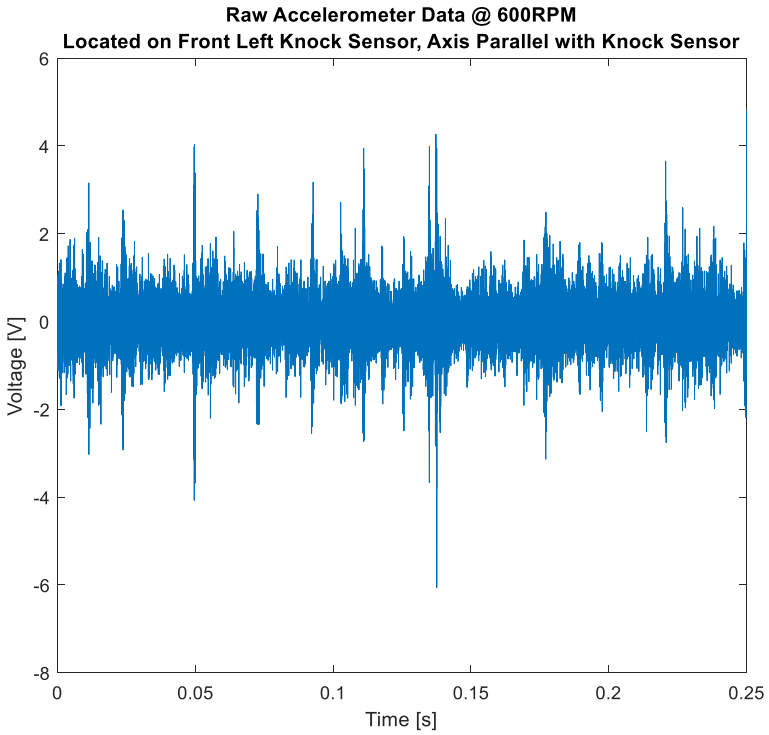


Figure 35: Raw accelerometer data from front left knock sensor

Encoder Data

Two signals were obtained from the encoder: encoder ticks and trigger signal.

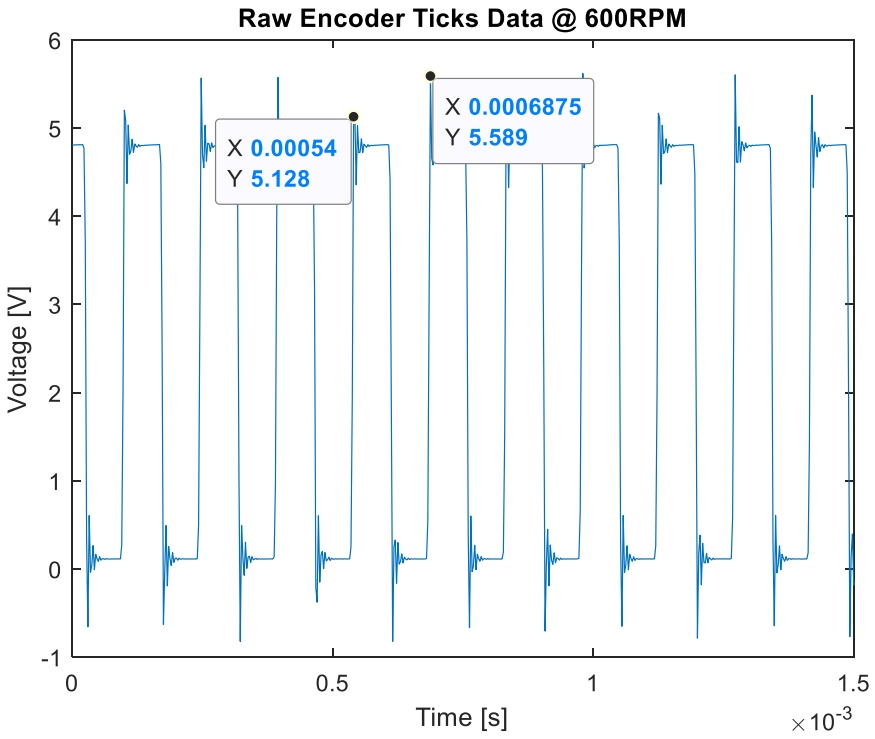


Figure 36: Raw encoder ticks @ 600rpm

Encoder ticks indicate the 0.5-degree travel of the crankshaft. Figure 36 is a plot of the raw encoder ticks data at an engine speed of 600 rpm. To verify the data, two points that indicate the start of an on switch on consecutive occasions are displayed. The time differences between these two points indicate the time it took for the crankshaft of the engine set to 600 rpm to travel 0.5 degrees. Assuming that the engine is running at a constant speed of 600 rpm the following calculation obtains the theoretical time difference between the two points selected in Figure 33. The difference between the theoretical and the actual time durations can be explained by the fact that actual engines are not capable of holding the speed constant. Therefore, the encoder ticks data obtained is considered to

accurately represent the change in the rotational position of the crankshaft with respect to time.

$$\Delta t_{theoretical} = 0.5° \cdot \frac{rev}{360°} \cdot \frac{min}{600\,rev} \cdot \frac{60\,s}{min} = 1.3889 \times 10^{-4} s \qquad (4.1)$$

$$\Delta t_{actual} = t2 - t1 = 0.0006875 - 0.00054 = 1.4750 \times 10^{-4} s \qquad (4.2)$$

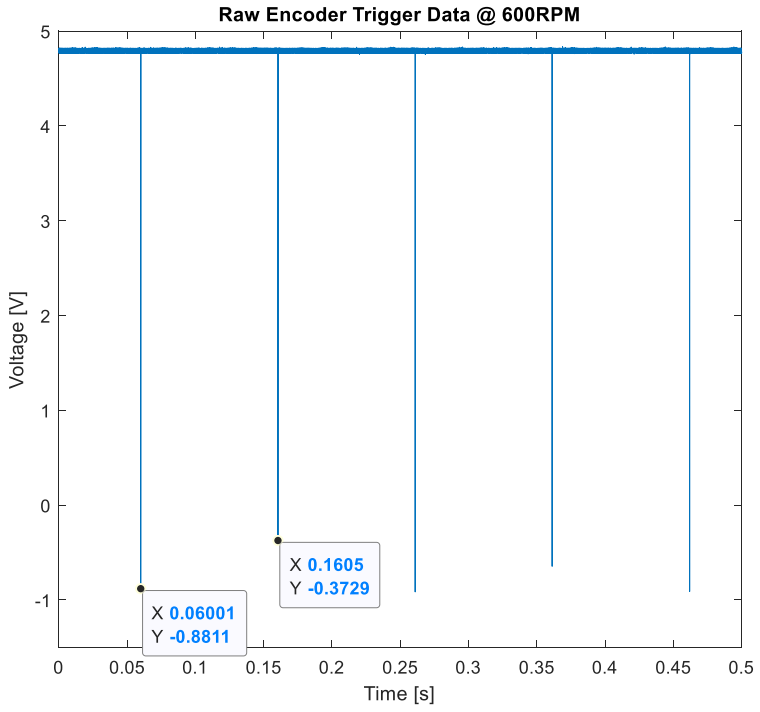


Figure 37: Raw encoder trigger @ 600rpm

Trigger signal indicates the occurrence of the Top-Dead-Center (TDC) in Cylinder 1, which occurs once per revolution of the crankshaft. Figure 37 is a plot of the raw encoder trigger data at an engine speed of 600 rpm. To verify the data, two points that indicate the start of an on switch on consecutive occasions are displayed as shown in Figure 34. With similar assumptions and method, the data is verified as shown in the calculations below.

With similar reasoning for the differences observed in the two time durations, the trigger data is considered to accurately represent the occurrence of TDC in Cylinder 1.

$$\Delta t_{theoretical} = 1\ rev \cdot \frac{min}{600\ rev} \cdot \frac{60\ s}{min} = 0.1000s \quad (4.3)$$

$$\Delta t_{actual} = t2 - t1 = 0.1605 - 0.06001 = 0.1005s \quad (4.4)$$

Pressure Data

Raw pressure data of Cylinder 1 is plotted in Figure 38. An overlay of pressure traces from all the cylinders are plotted in Figure 39 to validate the firing order of the engine. The order of the pressure peaks is shown to align with the firing order indicated in Section 4.1.1 which verifies that the pressure transducers were operating and placed properly.

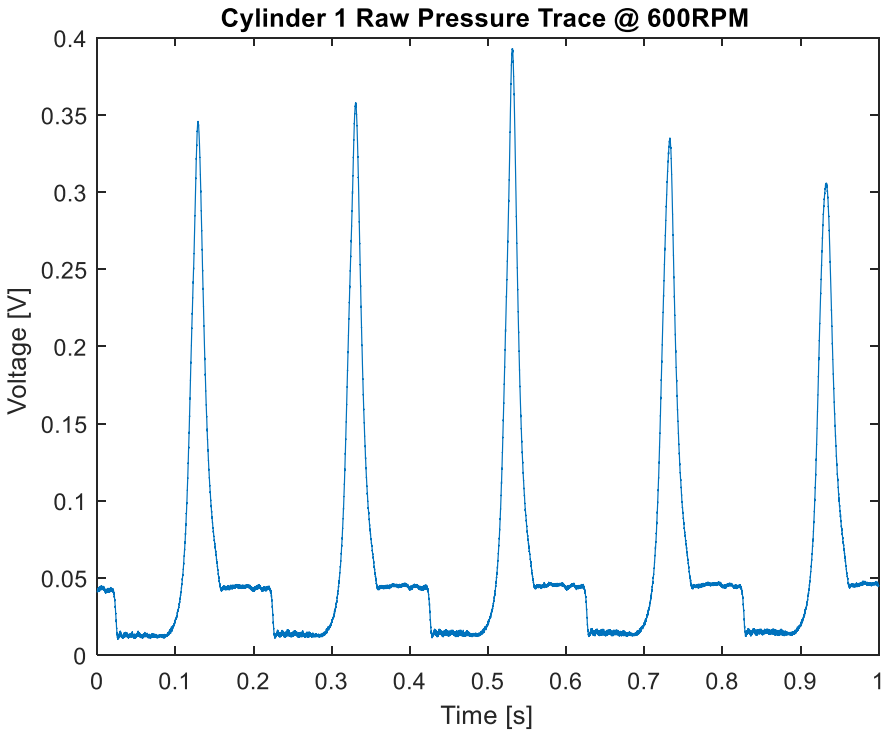


Figure 38: Raw Cylinder 1 Pressure Trace @600RPM

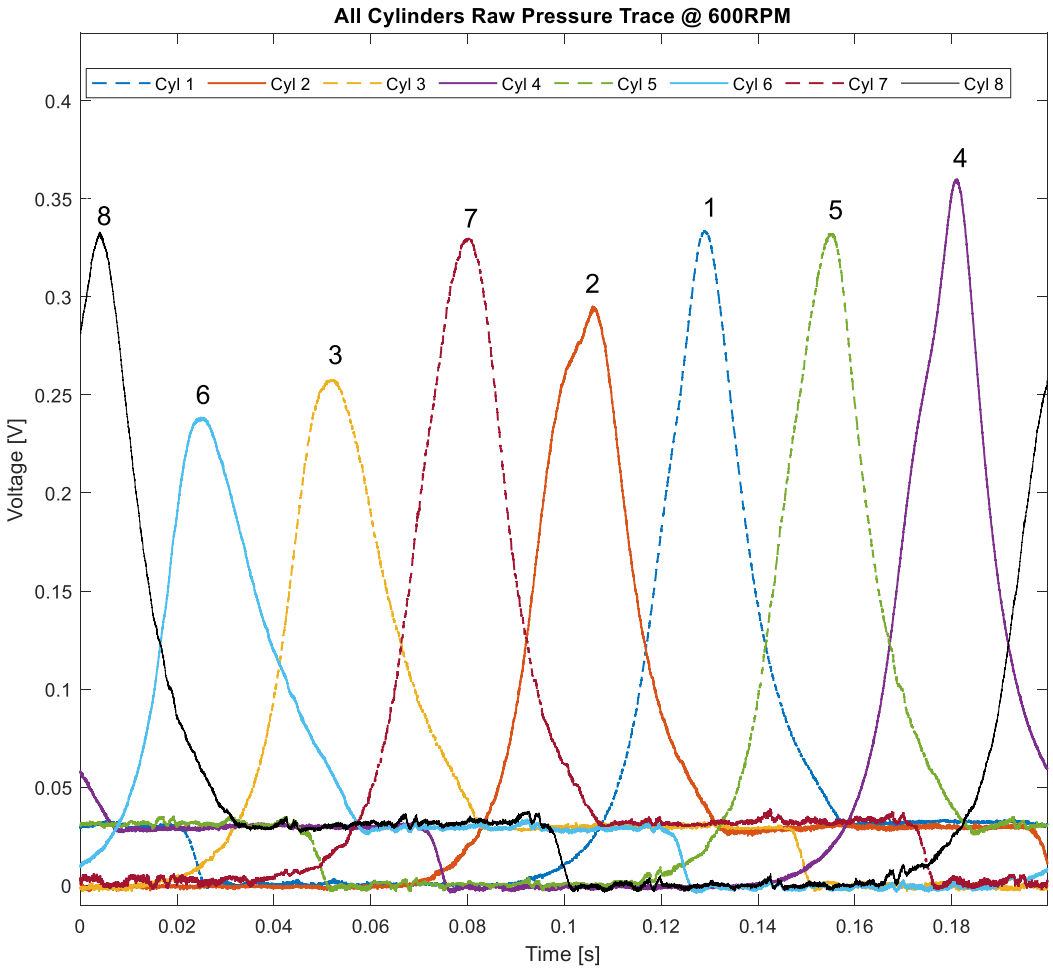


Figure 39: Overlay of all of the raw pressure traces from each cylinder @600RPM

4.3 Proposed Data Pre-Processing Strategy Implementation

In this section, the implementation of data pre-processing to transform the raw data into a compatible form for feature extraction is described in detail. As shown in Section 4.2.3, raw data consists of three different types of measurements; encoder, pressure and acceleration in the time domain. The final form required for feature extraction is acceleration measurement in the crank angle domain that is segmented into individual

cycles with corresponding fault class labels. The following sections describe in detail, the steps taken to transform the data.

Step 1: Filtering of Accelerometer Data

The accelerometer measurements were filtered using a Butterworth low-pass filter to remove noise. The sampling frequency of the vibration measurement was 100 kHz and the frequency response of the accelerometer is 1.2 Hz to 10 kHz. Therefore, the cutoff frequency was set to 10 kHz with a 12^{th} order filter with phase compensation to eliminate phase shift. Figure 40 shows the raw accelerometer signal and the filtered signal.

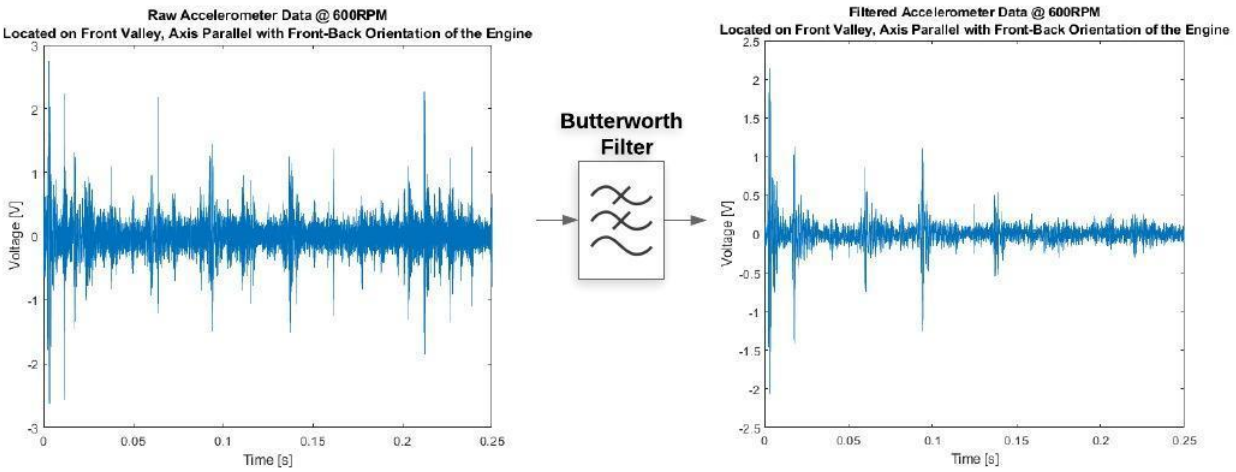


Figure 40: Filtering of accelerometer data using Butterworth Filter

Step 2: Encoder Measurement Analysis

The purpose of the encoder measurement analysis is to obtain crank angle versus time relationship and to obtain the occurrences of the Top-Dead-Center (TDC) of Cylinder1. The crank angle versus time relationship is utilized to resample the pressure and

acceleration measurements in crank angle domain. The timestamps at which TDC occurs in Cylinder 1 is utilized to segment the signals into aligned individual cycles.

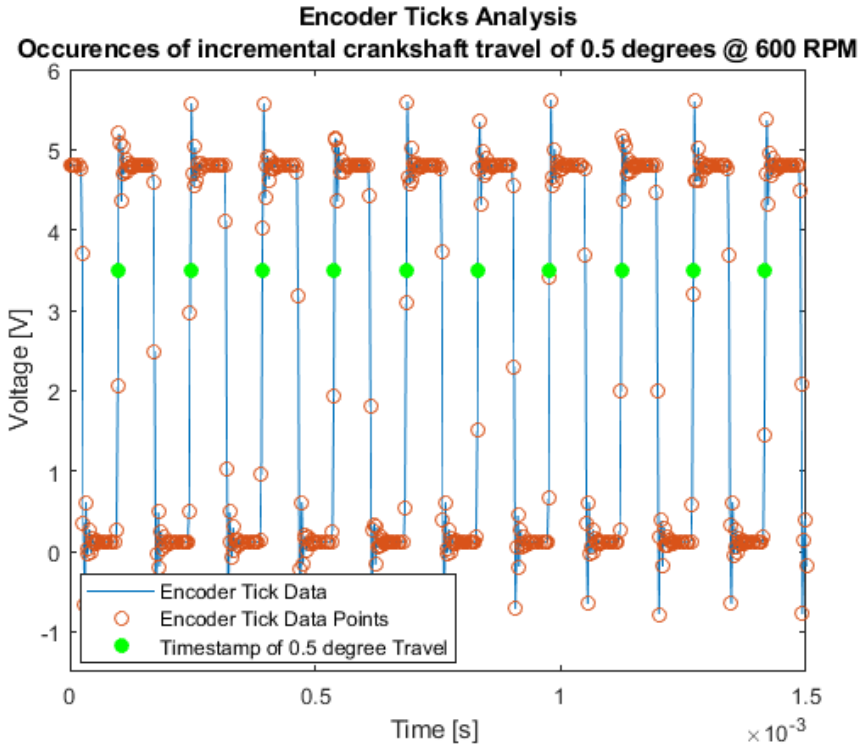


Figure 41: Encoder tick signal

The relationship between crank angle and time was determined by obtaining the timestamps of the instances where the light indicator switch of the encoder was flipped from “off” to “on”. More specifically, the instances at which the voltage was equal to the threshold value were obtained on the rising edge. The threshold value of 3.5 V was utilized as specified by the manufacturer. The exact timestamp was estimated through linear interpolation of two points immediately before and after the threshold value on the rising edge as shown in Figure 41.

Rotational speed of the crankshaft is derived from this crank position versus time relationship as shown in Figure 42 to validate the alignment of the set engine speed and the actual engine speed. It can be observed that the average engine speed aligns with the set speed of 600RPM. However, significant fluctuations also exist in the speed profile, which aligns with the unstable behavior described by Heywood at idle conditions [4]. The fluctuations in engine speed is a direct reflection of the fluctuations in combustion quality as mentioned by Heywood [4].

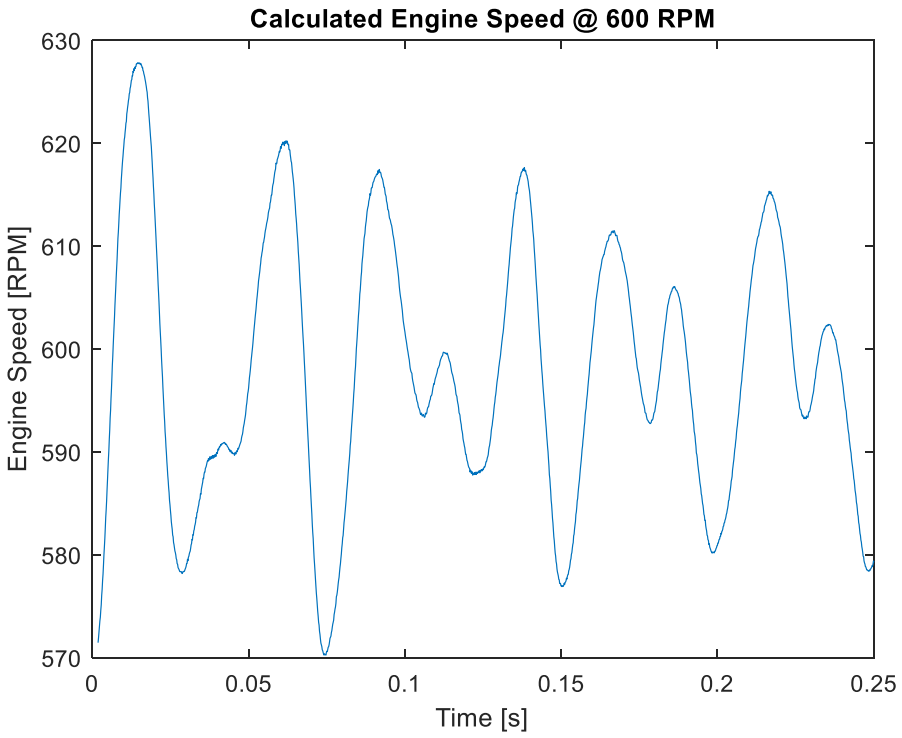


Figure 42: Calculated engine speed @ 600RPM

The occurrences of TDC in Cylinder 1 referenced in time were obtained using the trigger signal. A similar approach was used for encoder tick signal analysis to retrieve the

time at which the switch was flipped as an indicator for the occurrence of the TDC in Cylinder 1. It is important to note that TDC occurs twice in a given engine cycle as shown in Figure 43. Therefore, timestamps obtained by the trigger signal are not sufficient enough to synchronize individual cycles in the data set. Figure 43 illustrates that a piston of an engine reaches TDC once at the end of the compression stroke and again at the end of an exhaust stroke. With this knowledge, the cycle can be synchronized with additional information provided by Cylinder 1 pressure trace.

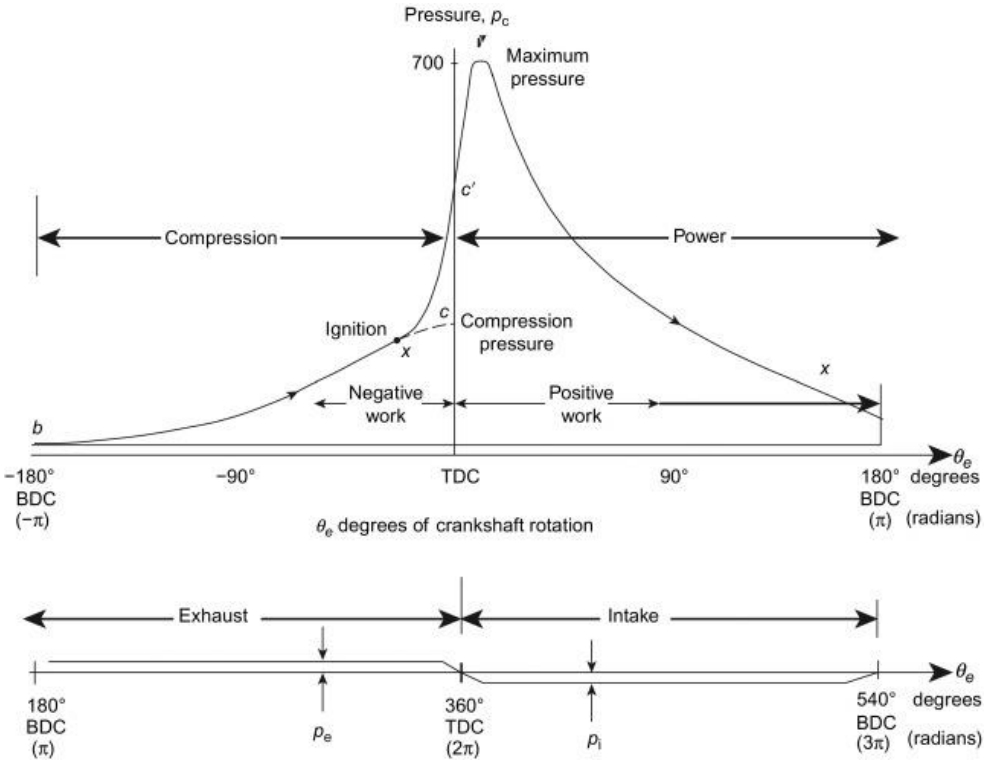


Figure 43: Pressure trace of a cylinder [62]

Figure 44 shows an overlay of trigger signal and Cylinder 1 pressure trace. Given the timestamp of a specific occurrence of TDC of Cylinder 1, pressure trace of Cylinder 1

can be referenced to see if this particular occurrence was at the end of a compression stoke or the exhaust stroke. If the pressure trace at the given timestamp is close to 0 bar, the current TDC can be concluded to have occurred upon an exhaust stroke. On the other hand, if the pressure value is well above 0, it can be concluded that the current TDC occurred upon a compression stroke. The timestamp associated with the exhaust stroke was chosen as the engine event at which the cycles will be segmented at. Therefore, the timestamps at which the TDC occurred at upon the exhaust stroke was obtained and was defined as the start timestamps indicating the start of a new cycle. It should be noted that there exists an offset between the occurrence of the TDC in Cylinder 1 presented by the trigger signal and the TDC that can be inferred by using the pressure signal. This offset is as expected as specified in Section 4.1.3.

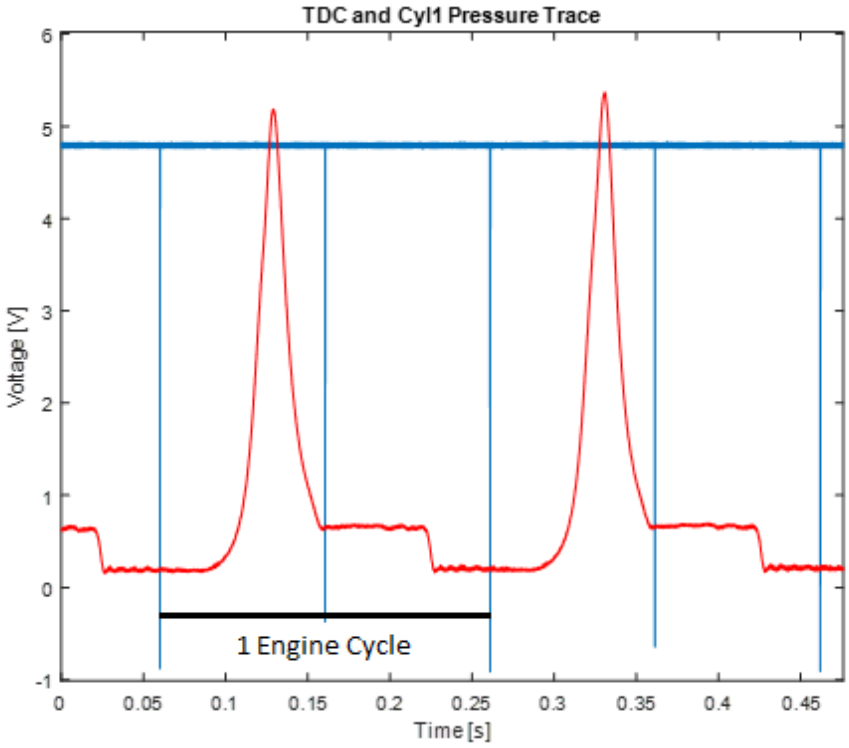


Figure 44: Overlay of trigger data and Cylinder 1 pressure trace @ 600RPM

Step 3: Resampling of Transducer Measurements

The crank position versus time relationship was used to resample the pressure and vibration measurements. The voltage of the transducer measurements at each crank angle – timestamps were obtained through linear interpolation. Figure 45 and Figure 46 show the plot of the transducer measurement in time and crank angle domain.

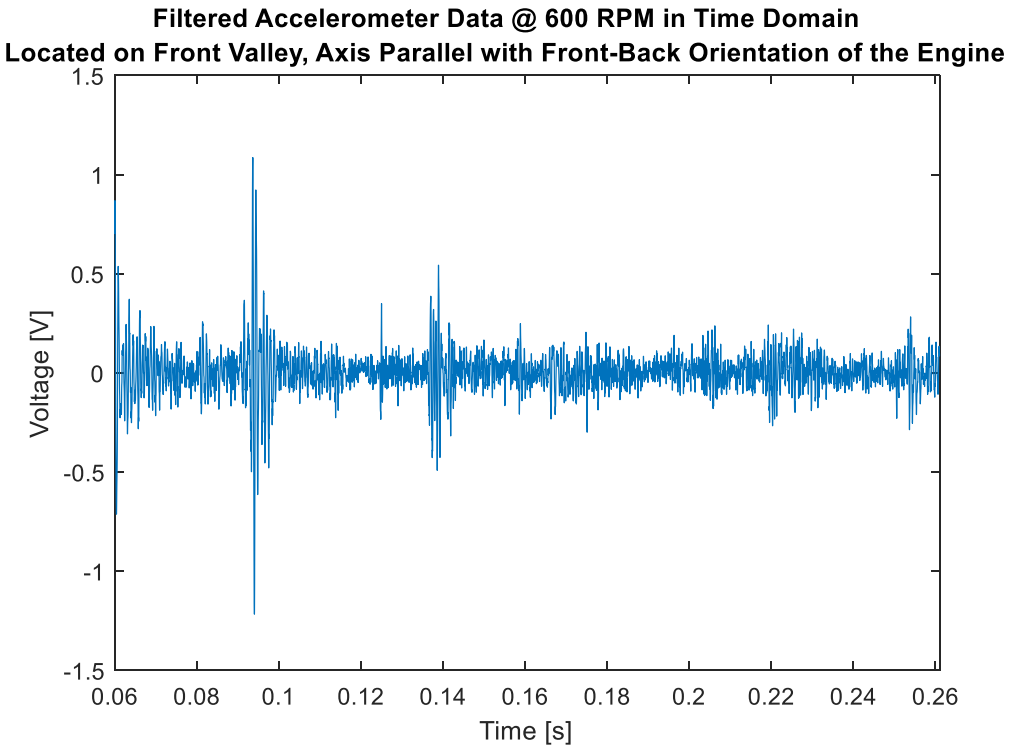


Figure 45: Filtered accelerometer data in time domain

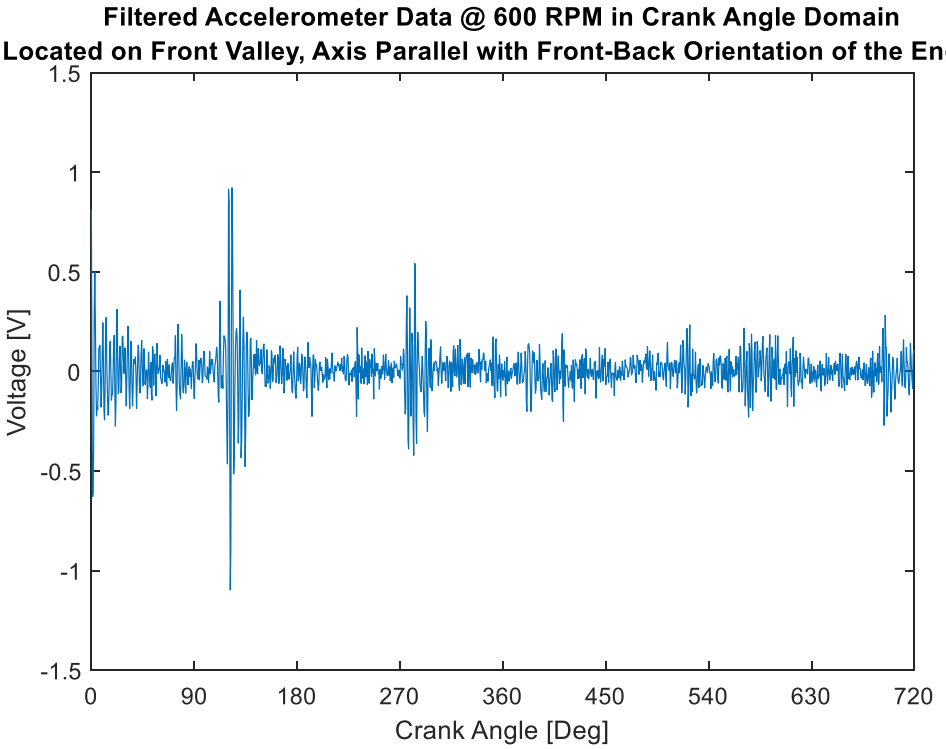


Figure 46: Filtered accelerometer data in crank angle domain

Step 4: Engine Cycle Segmentation

The start timestamps obtained in Section 4.3.2 is used to segment the transducer measurements into individual cycles. The start timestamps are referenced by a specific engine event to ensure the alignment of all the cycles obtained from different test files. The alignment is required as the data acquisition was randomly triggered which means that each test file starts at different positions of the crank angle that relate to different events of the engine. In this research, an engine cycle is defined as a 720-degree rotation of the crank angle that is aligned at 131.662 degree after the occurrence of TDC in Cylinder 1 upon an exhaust stroke event. Each measurement per cycle consists of 1440 points with a crank angle resolution of 0.5 degrees. Due to the fact that the offset of 131.662 degree is not

divisible by the crank angle resolution of 0.5 degrees, the start timestamp and the crank angle timestamp never shared a common timestamp. Therefore, first crank angle timestamp that occurs after the start timestamp is used as the start of the cycle. And the end of the cycle is marked by the first crank angle timestamp that occurs after next start timestamp.

Step 5: Data Labeling

Pressure trace provides very accurate and insightful information regarding the quality of the combustion occurring inside the cylinder. It would be ideal to develop a condition monitoring tool that utilizes in-cylinder pressure data from the engine. However, it is very unrealistic to have in-cylinder pressure transducers in operating vehicles due to their extremely high cost. Therefore, information obtained through pressure traces in test engines is correlated with more convenient and less costly vibration measurements. This correlation is performed through labeling the engine conditions of the vibration measurements using information extracted from the pressure traces.

For this experiment, the analysis of the pressure trace was performed by calculating the Indicated Mean Effective Pressure (IMEP). IMEP was calculated by integrating the pressure signal from 180-degree crank angle before TDC to 180-degree crank angle after TDC with respect to displacement volume as shown by the following equation, where $p(i)$ is the cylinder pressure at crank angle position i, $V(i)$ is the cylinder volume at crank angle position i, V_s is the cylinder swept volume, θ_0 is the bottom dead center (BDC) induction integer crank angle position, and θ_n is the BDC integer crank angle position [73].

$$IMEP = \frac{\Delta\theta}{V_s}\sum_{i=\theta_0}^{\theta_n} p(i)\frac{dV(i)}{d\theta} \tag{4.5}$$

If the IMEP value is below the specified lower limit, it is considered to be a misfire which is the fault condition of interest in this research.

Figure 47 shows a plot of all the pressure traces in one, previously defined, engine cycle. The plot shows that at a given engine cycle, a total of 360-degree pressure trace segment consisting of compression and power strokes required for IMEP calculation for every cylinder is not available within the specified engine cycle.

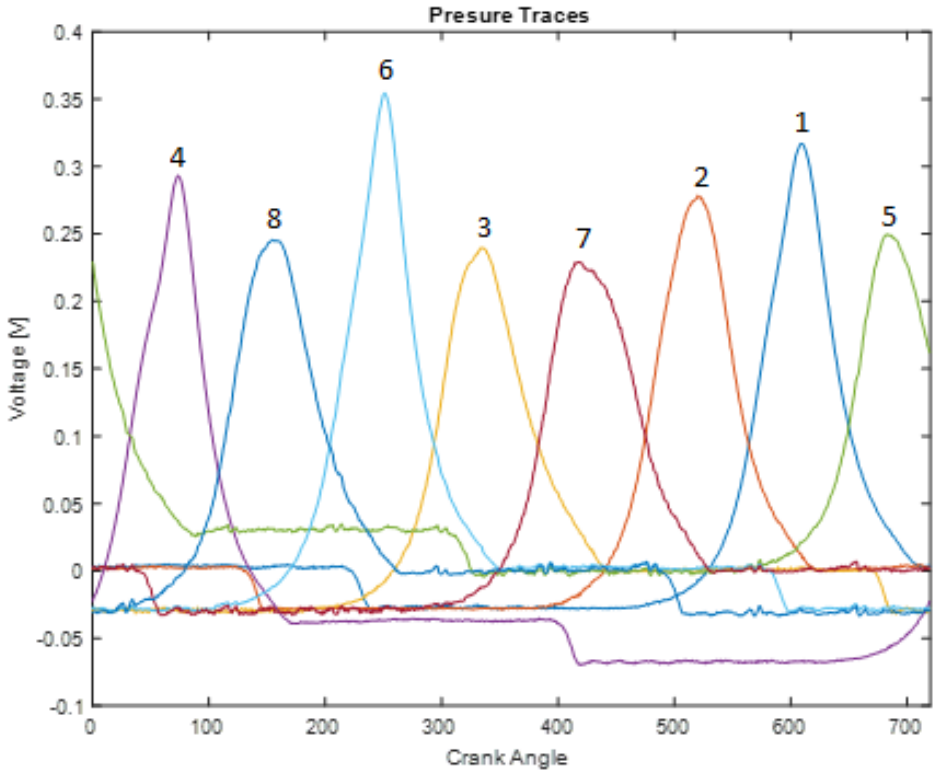


Figure 47: All pressure traces for 8 cylinder engine cycle

Figure 48 to Figure 50 are plots from two consecutive cycles for Cylinders 2, 5, and 8, respectively. Cylinder 2 is an example of those cylinders that have their compression and power strokes fully represented in the segmented engine cycle. On the other hand, Cylinders 5 and 8 are examples of those cylinders that do not have their compression and power strokes fully represented in the segmented engine cycle as shown in Figure 47. Also, it can be observed that the pressure trace of Cylinder 5 flows into the next engine cycle while the pressure trace of Cylinder 8 flows from the previous engine cycle. This infers that in order to calculate the IMEP of the combustion that is happening in Cylinder 5 for the specified engine cycle, the pressure trace for the following cycle is required. In a similar manner, the pressure trace of the previous cycle is required for IMEP calculation of the combustion happening in the Cylinder 8 for the same engine cycle. This concept was adopted for the IMEP calculations of those cylinders that were affected by the segmentation of the engine cycles. In addition, the very first and last cycles of every test were discarded as accurate calculation of the IMEP values for all of the cylinder were not obtainable as full compression and power strokes are not available.

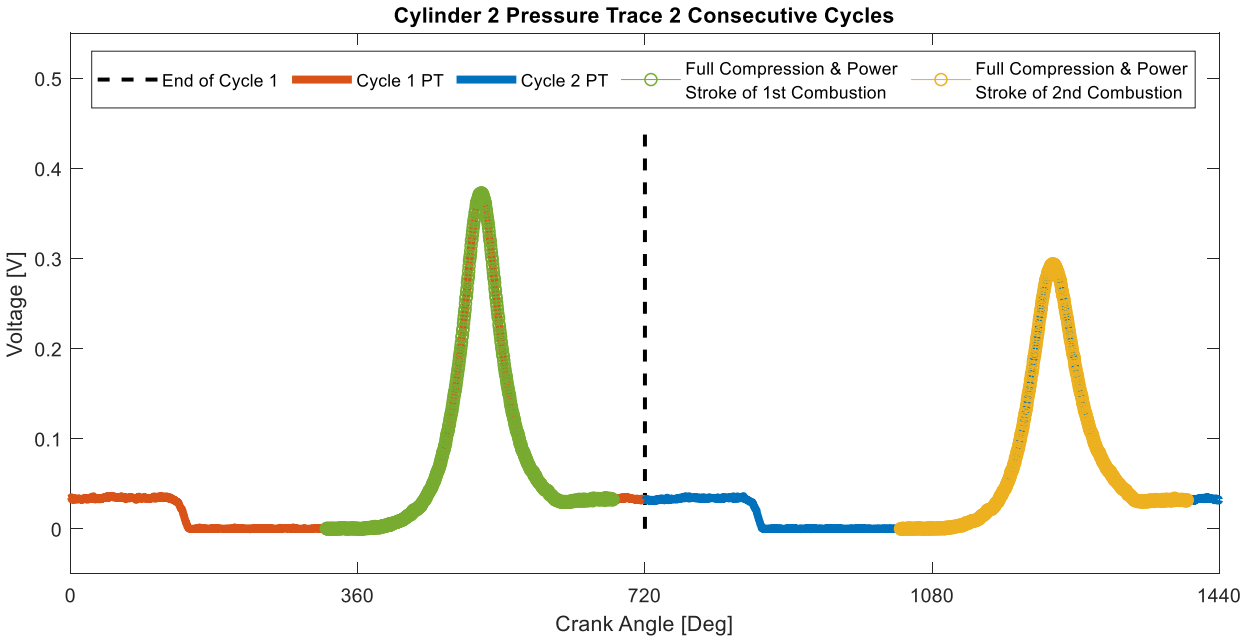


Figure 48: Cylinder 2 Pressure Trace of 2 Consecutive Cycles

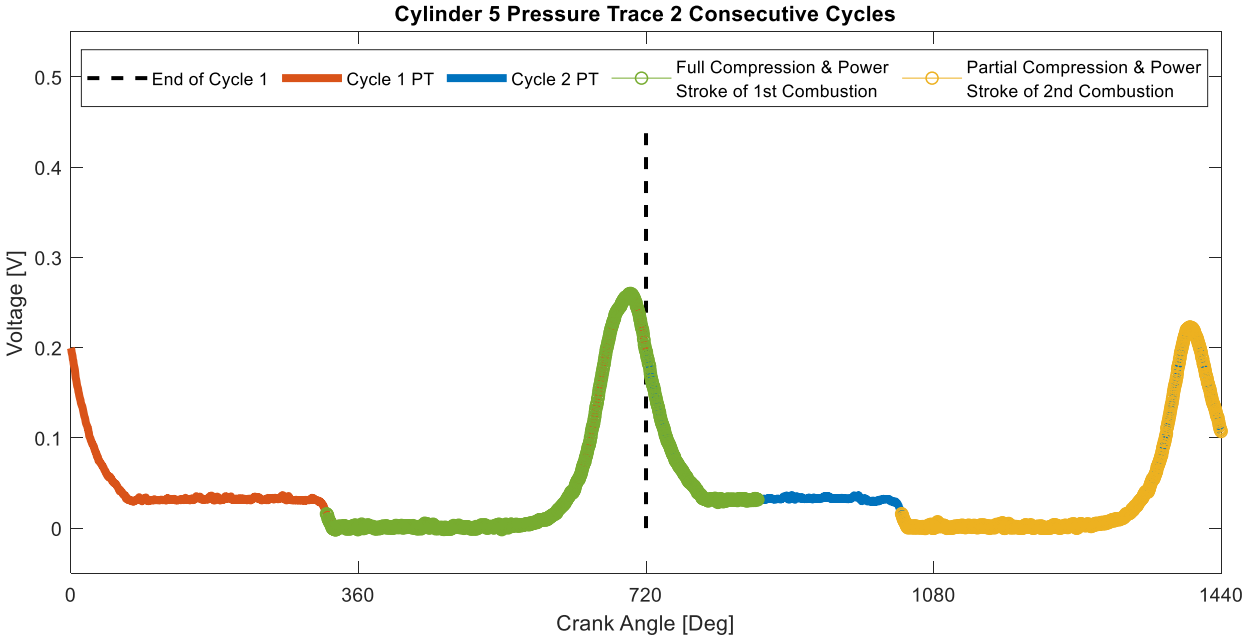


Figure 49: Cylinder 5 Pressure Trace of 2 Consecutive Cycles

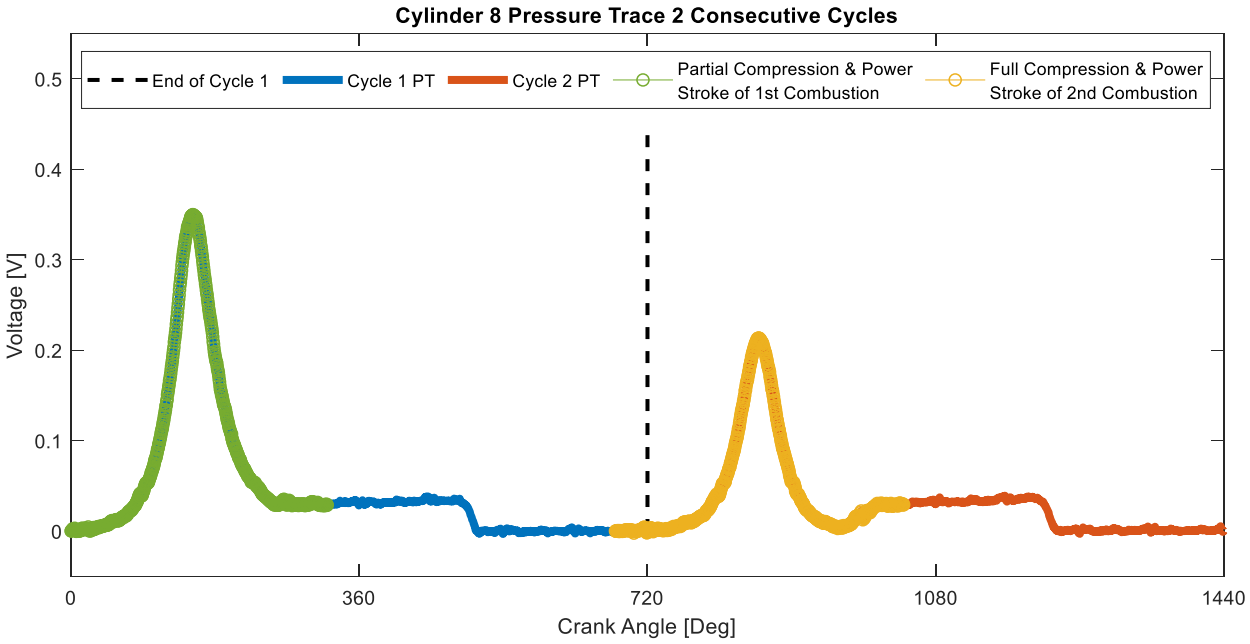


Figure 50: Cylinder 8 Pressure Trace of 2 Consecutive Cycles

It should be noted that the vibration measurements were segmented without carrying over the vibrations that occurred during the full compression and power strokes of the cylinders. The vibration measurements were not carried over due to the definition of an engine cycle that was established earlier. This lack of data representation may lead to difficulty in correlating the pressure to the vibration relationship of faults occurring in the affected cylinders.

Pre-Feature Extraction Data Summary

The following table summarizes the labeled cycles in the dataset obtained through engine testing. It is very important to note the imbalance of the different conditions in the data collected. This is primarily due to the way the faults were simulated. In previous research conducted, misfire faults were simulated by disconnecting the spark plug which

ensures that every cycle of the specified cylinder misfires [22]. This method of simulating the fault allows for a balanced distribution of the data to be obtained. However, in this research the faults were induced by simulating a condition of the engine where the probability of incomplete combustion is high. Therefore, the occurrence of faults or the distribution of faults were not guaranteed.

Table 2: Summary of labeled data

600 RPM		Cycles
Healthy		8002
Fault 1	Misfire in Cylinder 1	66
Fault 2	Misfire in Cylinder 2	9
Fault 3	Misfire in Cylinder 3	188
Fault 4	Misfire in Cylinder 4	7
Fault 5	Misfire in Cylinder 5	369
Fault 6	Misfire in Cylinder 6	81
Fault 7	Misfire in Cylinder 7	167
Fault 8	Misfire in Cylinder 7	76

Three sample cycles that consist of one faulty cycle and two different healthy cycles that were labeled in the previous section using IMEP calculations are plotted in the following series of figures. The fault class of the faulty cycle is 1 which indicates that the fault has occurred in Cylinder 1. Figure 51 is an overlay of pressure traces from Cylinder 1 for three different cycles. Figure 52 to Figure 54 are the vibration data from the three-axis accelerometer located on the front valley. Figure 55 is the vibration data from the accelerometer located on the front left knock sensor.

When observing the pressure signals, the maximum pressure level and the area under the pressure trace of faulty cycle is noticeably smaller than that of the healthy cycles which would numerically be translated into the IMEP values. Therefore, it can be concluded that faulty cycles are easily identifiable amongst a mixture of healthy and faulty cycles for a given cylinder.

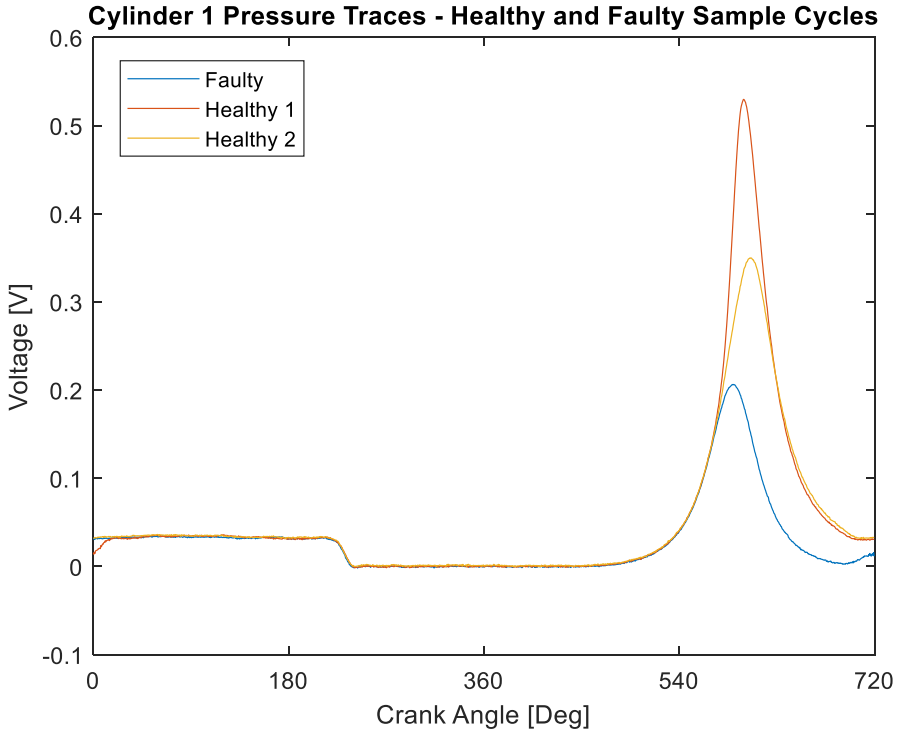


Figure 51: Cylinder 1 Pressure Traces for Faulty and Healthy sample cycles

When observing all the vibration data, the difference between healthy and faulty data is not as clear. The distinguishable characteristics such as big spikes and locations of those big spikes do not seem to be unique for healthy cycles. When observing the spikes, it can be seen that there does not seem to be a commonality in the locations of the spike

with respect to crank angle. In terms of amplitude, it can be noticed that Healthy 2 cycle shows lower amplitude than the Faulty cycle in all of the vibration plots. Therefore, it can be concluded that faulty cycles are not easily identifiable amongst a mixture of healthy and faulty cycles without further analysis or without advanced FDD tools.

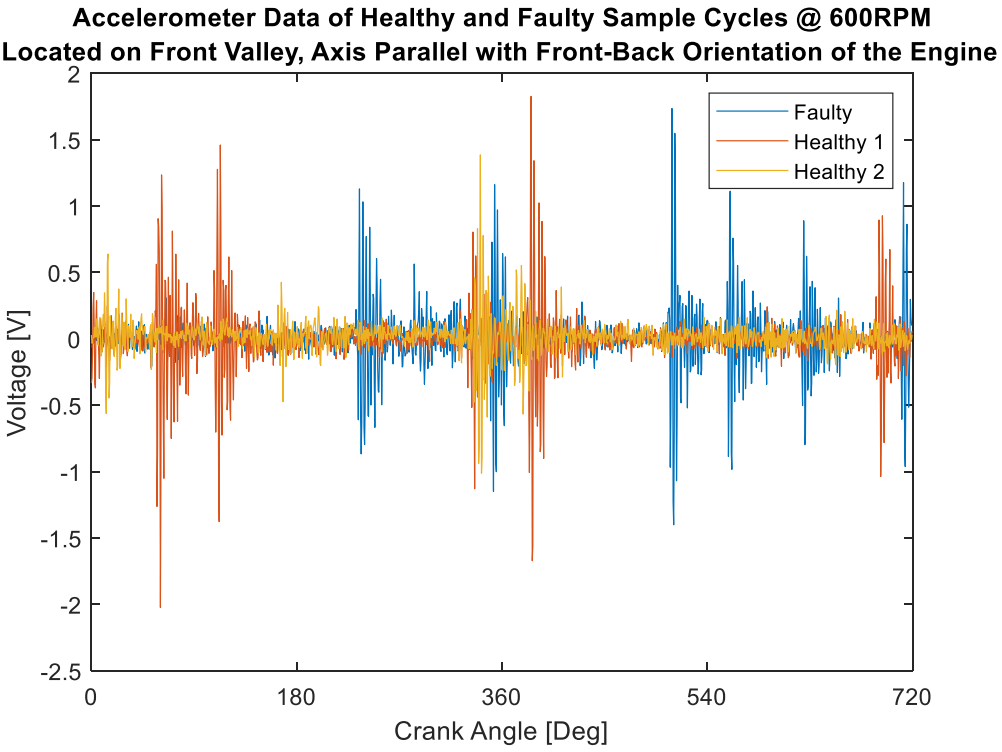


Figure 52: Front-Back Axis Accelerometer data from Front Valley for Faulty and Healthy sample cycles

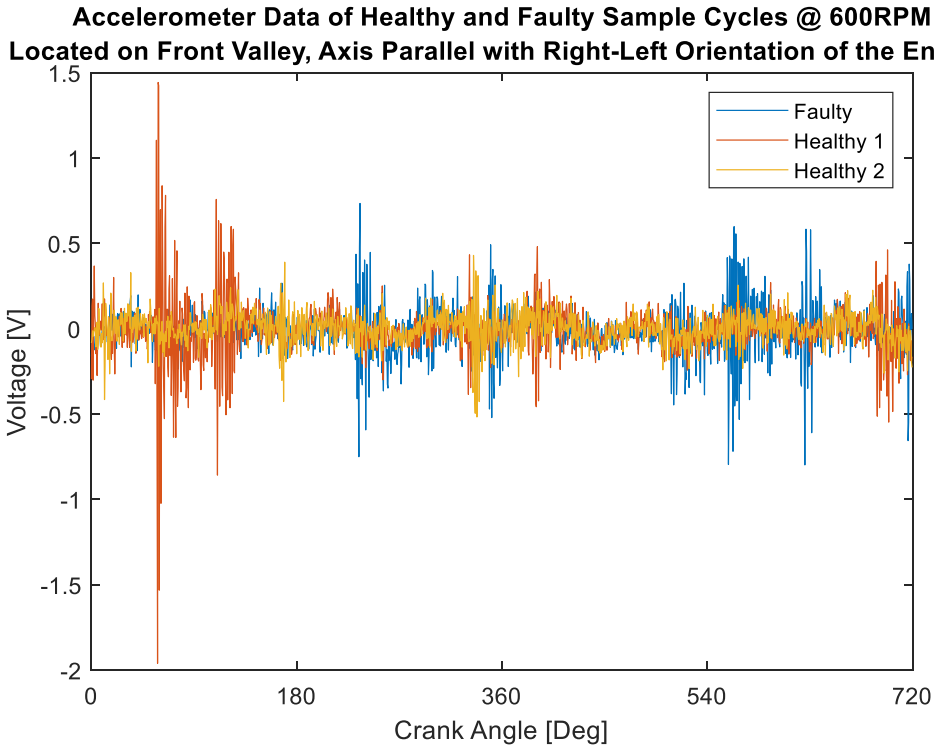


Figure 53: Right-Left Axis Accelerometer data from Front Valley for Faulty and Healthy sample cycles

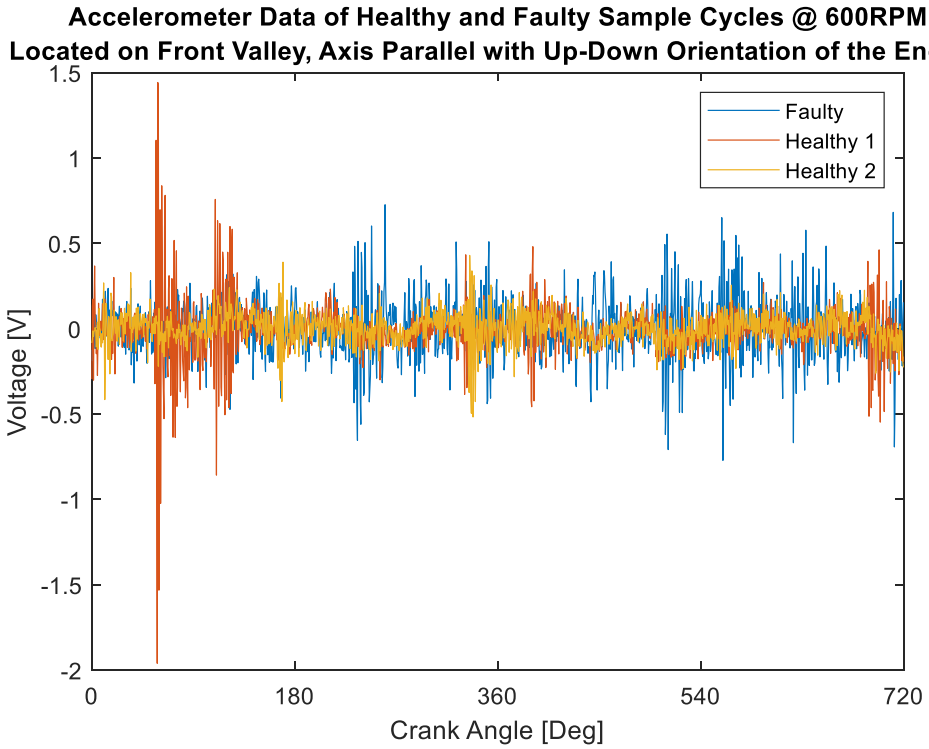


Figure 54: Up-Down Axis Accelerometer data from Front Valley for Faulty and Healthy sample cycles

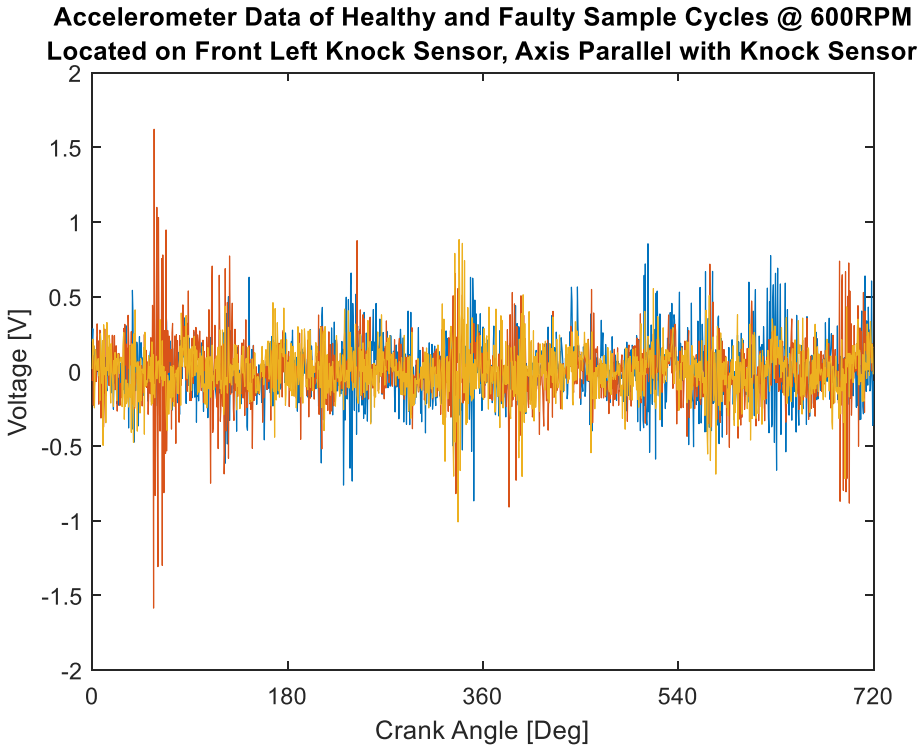


Figure 55: Axis Accelerometer data from Front Left Knock Sensor for Faulty and Healthy sample cycles

Chapter 5: Proposed FDD Strategy Implementation and Results

5.1 Proposed Feature Extraction Strategy Implementation

In this section, the implementation of the feature extraction strategy is described in detail. The purpose of the feature extraction strategy is to transform the pre-processed data, represented in Section 4.3, into features. The theory pertaining to the proposed strategy was provided in Chapter 3. This section focuses on elaborating on the parameters chosen to implement the strategy.

The proposed feature extraction method takes the user-defined baseline data and faulty data and generates features that represent how different the faulty data is in comparison to the baseline data. The proposed strategy utilizes the Extended-MSPCA to decompose both the baseline and faulty data through wavelet packet transformation, then transforms them into main components using PCA and generates features using statistical indices. The implementation of the strategy consists of a selection of wavelet function, wavelet level, and baseline data.

In this research, the algorithm is tasked to differentiate the misfiring cycle from the healthy data and differentiate the locations of the misfire. The difference between a misfire

occurring in one cylinder from the other is hypothesized to primarily depend on the occurrence of the misfire with respect to the crank angle domain. This hypothesis assumes that the differentiating characteristics of the misfire condition in comparison to the healthy condition will occur within the compression and power strokes of a given cylinder. Depending on the firing order for a given engine, each cylinder will fire at a different crank angle range. This infers that the wavelet function must have enough localization in the crank angle to be able to differentiate the same frequency components occurring 90 degrees apart. On the other hand, the differentiation of faulty and healthy cycles may be highly dependent on frequency. Therefore, a balance of frequency and time resolution must be achieved for the success of accurate characterization of faults.

Step 1: Wavelet Function Selection

A systematic process of selecting an optimal wavelet function for a given application does not exist. Consequently, the wavelet function that behaves similarly to the signal of interest is advised to be utilized for the wavelet transform. Therefore, in the case of fault detection and diagnosis, the ideal wavelet function to be utilized would be one that resembles the faulty components of the signal. However, the faulty components are not easily identifiable in the signals obtained, given the resources and time allowed for further analysis of the signals. Therefore, a wavelet family called Daubechies, which is commonly used for vibration in FDD applications, is utilized in this research. Daubechies wavelets have been utilized in the previous studies conducted by Ismail, Doghri, and Narendiranath for FDD applications on alternators and starters, internal combustion engines, and journal bearings, respectively with successful outcomes [12], [22], [74].

Daubechies wavelets are said to be compactly supported, which means that it will provide results that are more localized as opposed to global results. Daubechies wavelets consist of numerous individual wavelet functions that are differentiated by its number of vanishing moments. Wavelet db4 represents Daubechies wavelet with vanishing moments of 4. The increase in vanishing moments implies that more complex behaviors of the signal will be captured. In addition, Figure 56 shows that higher-order mother wavelets provide superior frequency resolution with sharper roll-off. However, the improvements made on frequency resolution results in a reduction in time resolution.

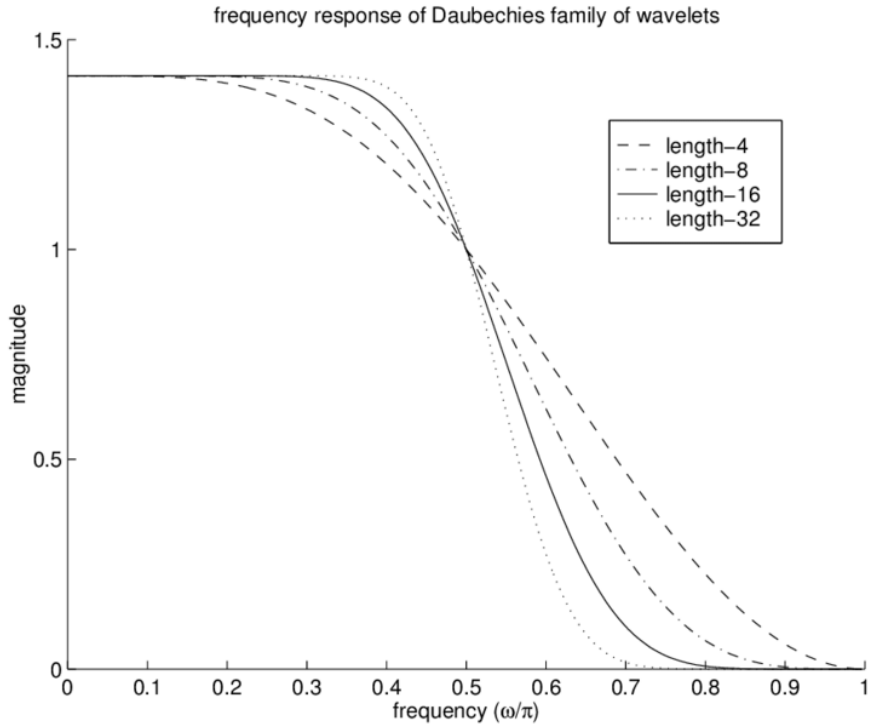


Figure 56: Frequency response of Daubechies wavelets [75]

For the feature extraction process component of this research, different wavelet functions are utilized to study their effects on the performance of the FDD strategy. These

wavelet functions consist of db4, db10, and db16 were utilized. Figure 57, 55, and 56 show the corresponding db16, db10, and db4 low pass and high pass filters that were utilized during Extended-MSPCA.

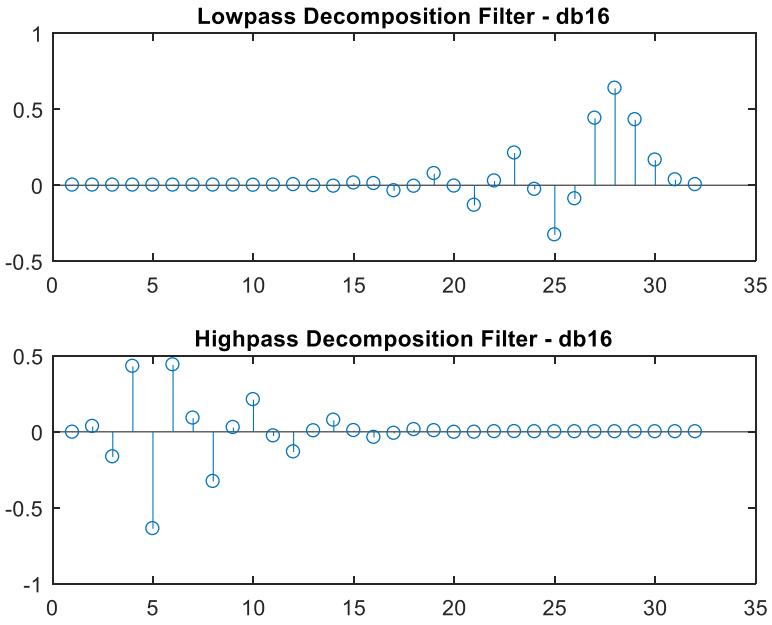


Figure 57: Low and high pass filter for db16

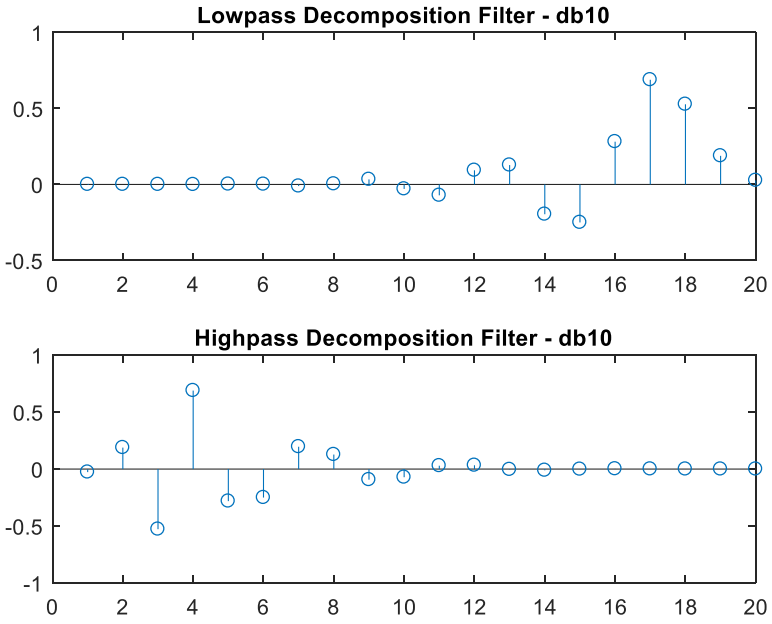


Figure 58: Low pass and high pass filter for db10

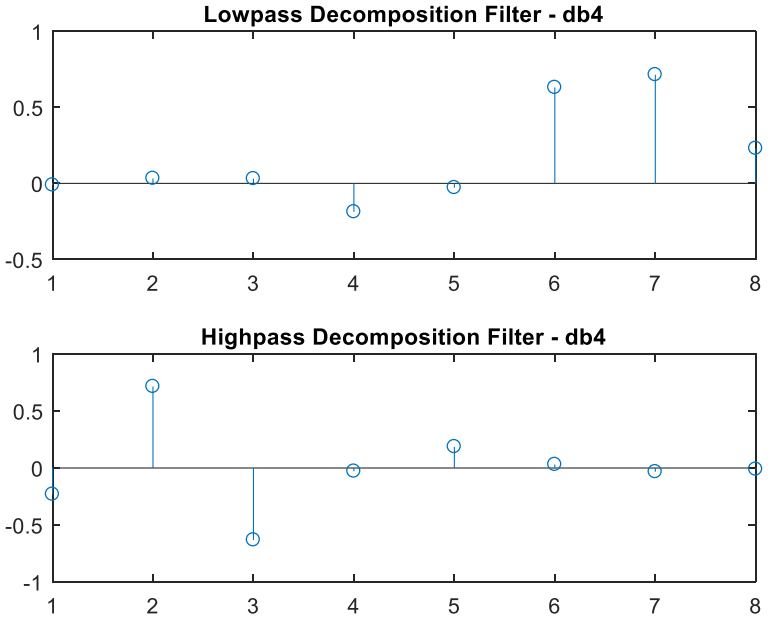


Figure 59: Low pass and high pass filter for db4

Step 2: Wavelet Level Selection

Wavelet levels dictate the number of frequency bins the given signal will be decomposed to. As wavelet levels increase, the number of frequency bins increases while the range of each frequency is decreased. In other words, the increase in wavelet levels leads to finer frequency resolution information. The increase in frequency resolution results in decreased crank angle resolution as crank angle resolution is halved at each subsequent decomposition level. The crank angle information is predicted to be crucial in the characterization of faults in different cylinders due to the cyclic operation of the engine. In addition, there is a maximum number of wavelet levels that can be used for a given sample signal and the wavelet being used for the analysis. The maximum number of wavelet levels can be obtained through the following expression:

$$L_{max} < log2\left(\frac{N_{sample}}{N_{wavelet}} - 1\right) \tag{5.1}$$

Given this relationship, the maximum wavelet level for each wavelet function is obtained and summarized in the table below. For this research, a wavelet level of 4 will be utilized for all three wavelets which is limited by db16's maximum wavelet level. The maximum level of wavelet is restricted in this research due to the short length of the dataset. In previous studies performed by Ismail and Doghri, longer sample lengths were utilized, which meant that higher wavelet levels could be utilized. For cycle-by-cycle monitoring, the only method of increasing the sample length is through the increase in angular resolution of the dataset.

Table 3: Max wavelet level for corresponding wavelet

Wavelet	Max Level
db16	5
db10	6
db4	7

Step 3: Defining Baseline

As discussed in Chapter 3, it was proposed that multiple baseline data approach is to be taken as opposed to the single baseline data approach. For the application of this strategy, the method of selection and the method of determining the number of baselines required are important components of the feature extraction strategy that must be considered further for a thorough study of the feature extraction strategy. However, in this research, a simplified method of selection is pursued for the interest of time. The simplified method of selection involves randomly selecting a specified number of individual cycles from all the healthy data set that have been collected. In this experiment, 10 different baseline data will be utilized for the feature extraction component.

In the previous study performed by Doghri where the FDD strategy was applied to detect constant fault conditions, averaging method of features that belong to the same fault condition was proposed [22]. The constant fault of an internal combustion engine can be defined as a condition that is known to manifest regardless of the cycle-to-cycle variations. The purpose of the moving average is to capture the dominant characteristics of a given fault signature. However, the application of a moving average is not recommended for

cycle-by-cycle monitoring, which is required for the detection of intermittent faults. This is due to the implementation methods of the developed FDD strategy and the type of fault the strategy is aimed to monitor. If averaging was performed during the development of the FDD strategy, averaging would also be required to guarantee the success of the classifier. If the unknown measurements are not averaged, the measurement will consist of noise and other cycle-specific variations that the training algorithm has not seen during the training process. The presence of noise and cycle-to-cycle variation of the unknown measurement may affect the results of the classification tool. Therefore, implementation methods must be taken into consideration during the development process of the FDD strategy.

Feature Summary

From all the cycles obtained and labeled, 100 randomly selected healthy cycles and the rest of the faulty cycles were defined as a finalized dataset, which consists of 1063 cycles, which equates to 10630 features. The number of healthy cycles was reduced to 100 cycles to even out the imbalance of data. Aside from 100 healthy cycles, 10 additional healthy cycles were randomly selected to form the baseline dataset. From the dataset of all the cycles, ~70%, ~20%, and ~10% of each class was randomly selected to form the training, validation and testing subsets. The testing subset was formed by randomly selecting ~10% of the total cycles in the dataset. For example, for the healthy condition there are a total of 100 cycles which equate to 1000 features. To select the testing set, 10 cycles from 100 cycles were randomly chosen and the testing set was formed by 100 features that have been generated from the 10 cycles that were selected. However, to divide

the remaining data into training and validation, 70% of the features were randomly chosen per condition to form the training set and the remainder formed the validation set. The breakdown of total cycles and features are shown in Table 4. The method chosen to divide the total data into subsets were chosen to ensure that the testing set will be executed in the way that the classifier is to be used in real application. Further explanations of each dataset and its importance will be elaborated in the classifier section.

Table 4: Data breakdown

	Cycles / Features	Training	Validation	Testing
Healthy	100 / 1000	- / 630	- / 180	90 / 900
Fault 1	66 / 600	- / 460	- / 130	7 / 70
Fault 2	9 / 90	- / 60	- / 20	1 / 10
Fault 3	188 / 1880	- / 131	- / 380	19 / 190
Fault 4	7 / 70	- / 50	- / 10	1 / 10
Fault 5	369 / 3690	- / 2580	- / 740	- / 370
Fault 6	81 / 810	- / 570	- / 160	8 / 80
Fault 7	167 / 1670	- / 1170	- / 330	17 / 170
Fault 8	76 / 760	- / 530	15 / 150	8 / 80
Total	1063 / 10630	- / 7430	- / 2120	108 / 1080

The features extracted consist of 16 coefficients for each sensor signal utilized, which results in a feature dimension of 16 x 8 or a total of 128 coefficients. This is a great reduction in dimensionality from the raw signal, which consists of 1440 data points for each sensor signal, resulting in 11520 data points. The reduction in dimensionality will reduce the time and computational resources required for the classification algorithm.

Figure 60 and Figure 61 show the features that were extracted from accelerometers on the valley and knock sensors, respectively. A general trend can be observed from the

features extracted from the valley accelerometers across all the fault conditions, including the healthy condition, and subtle differences that can be observed from one condition to the other. However, these differences are not easily translated into quantifiable relationships to characterize each fault condition. The difficulties found in differentiating each fault class from the visual representations of the feature sets validate the need for a sophisticated classifier algorithm that can infer subtle nuances of the features that are not easily identifiable.

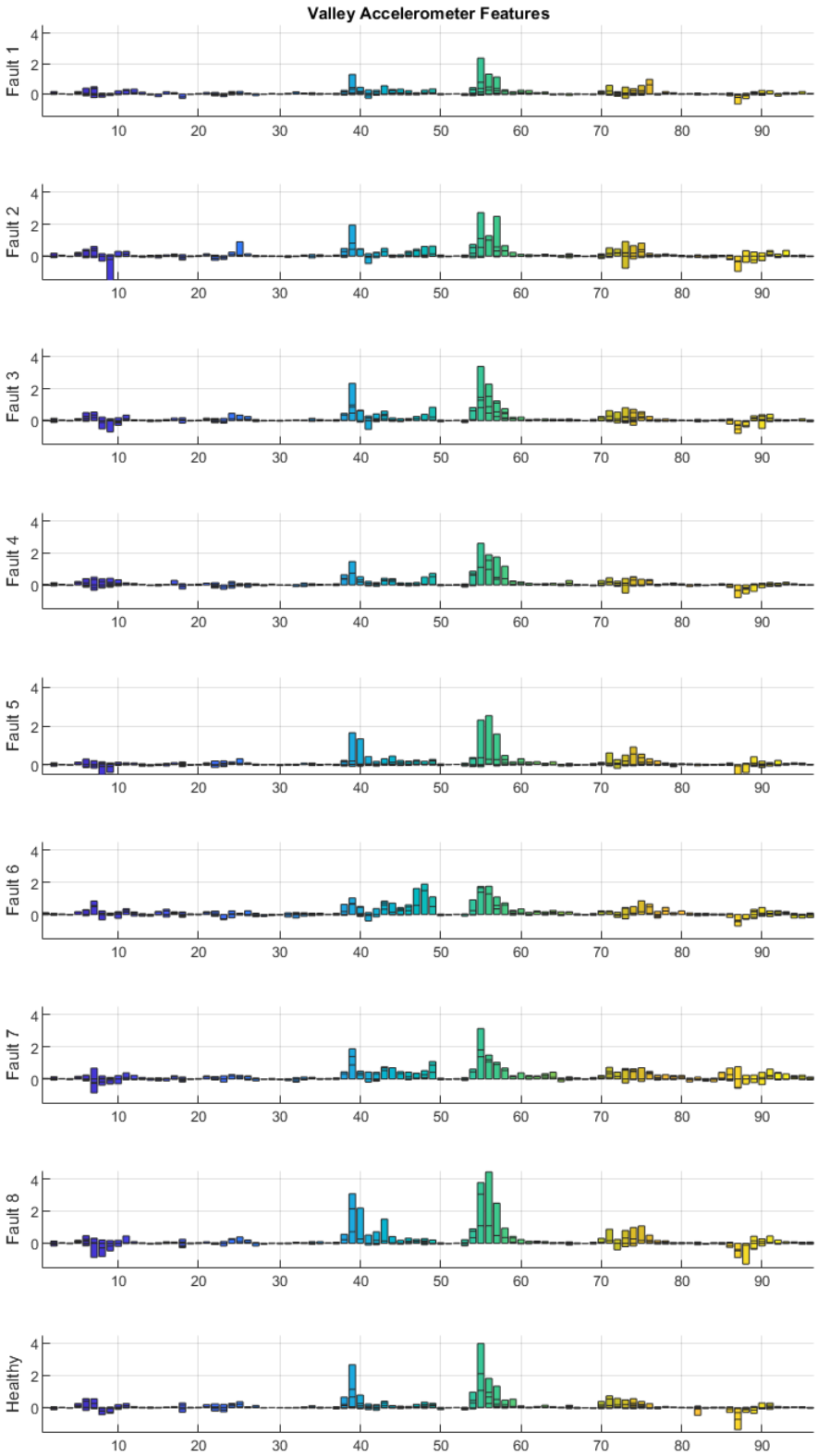


Figure 60: Features extracted from valley accelerometers

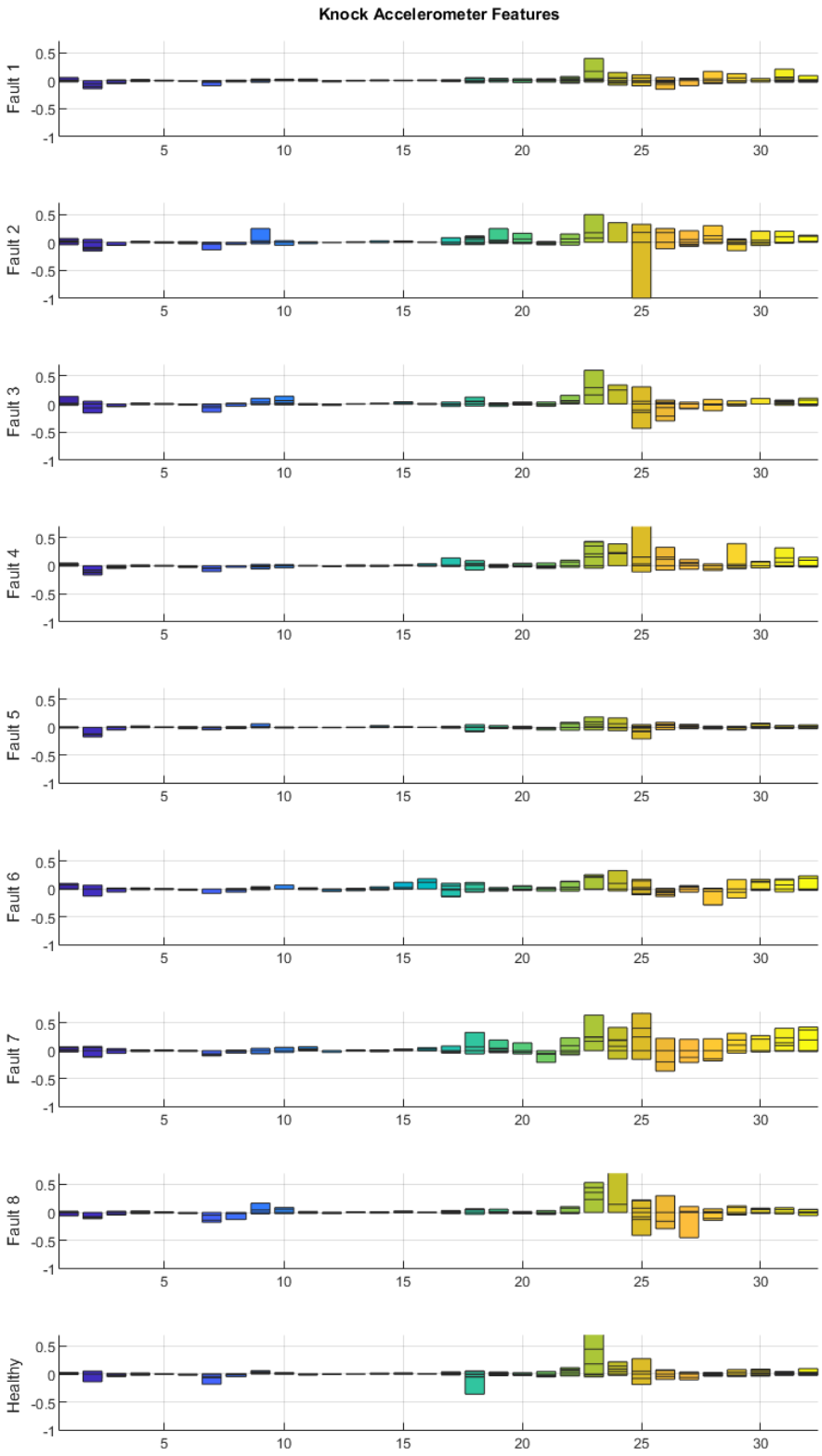


Figure 61: Features extracted from knock accelerometers

The table below summarizes the time it took for the generation of features for each wavelet type and wavelet level used. The time of feature extraction is very important for both the training and utilization of the FDD strategy. The Extended-MSPCA algorithm used in this research is sufficient for the purposes of evaluating the analysis method and its ability to perform analysis on a cycle by cycle basis. However, the algorithm will require extensive improvement in order for it to be capable of monitoring the engine on a cycle by cycle basis.

Table 5: Time duration for each feature extraction

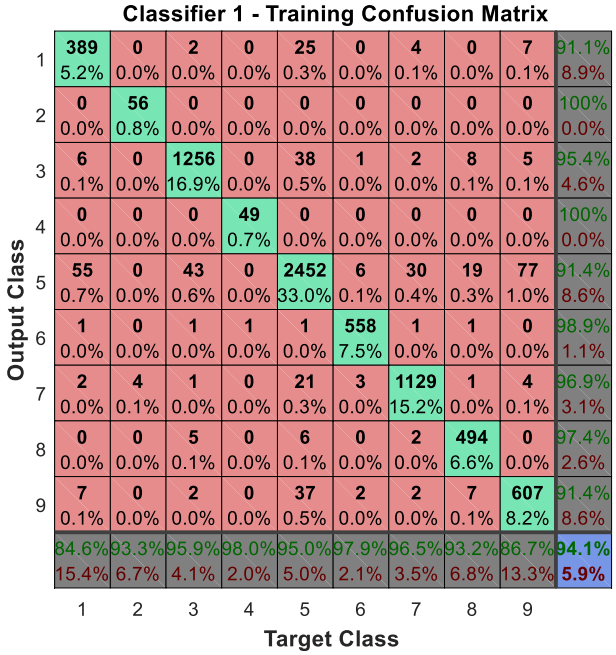
Wavelet Type	Wavelet Level	Time per Cycle/Baseline
db4	4	1.031360s
db10	4	1.199280s
db16	4	2.161937s

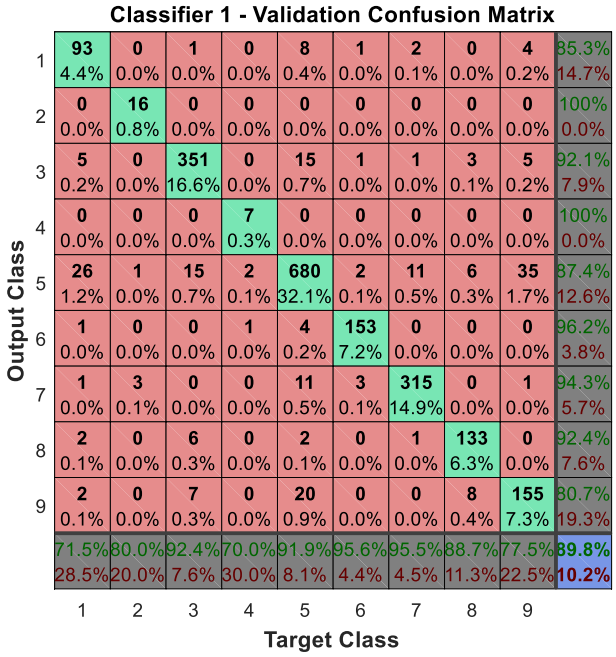
5.2 Results and Observations

In this section, the results of the developed FDD strategy are presented. The results will review the performances of three different classifiers, Classifier 1, 2 and 3, that have been developed using features extracted by three different wavelet functions, db4, db10, and db16. The hyper parameters such as number of hidden nodes, learning rate and stopping criteria, were tuned individually for each classifier to yield the most optimal classifier for the given dataset. For each classifier, the results are presented through four different confusion matrices. These include the training confusion matrix, validation confusion matrix, testing confusion matrix, and detection confusion matrix as shown in Figure 59 – 61.

The training confusion matrix shows how well the classifier was able to learn from the training data set. The validation confusion matrix shows how well the classifier is able to generalize. The testing confusion matrix shows the classifier's ability to classify the features that the classifier has not seen during training. And the detection confusion matrix shows the ability to correctly classify the given cycle using majority voting method.

Classifier 1: Wavelet Function db4





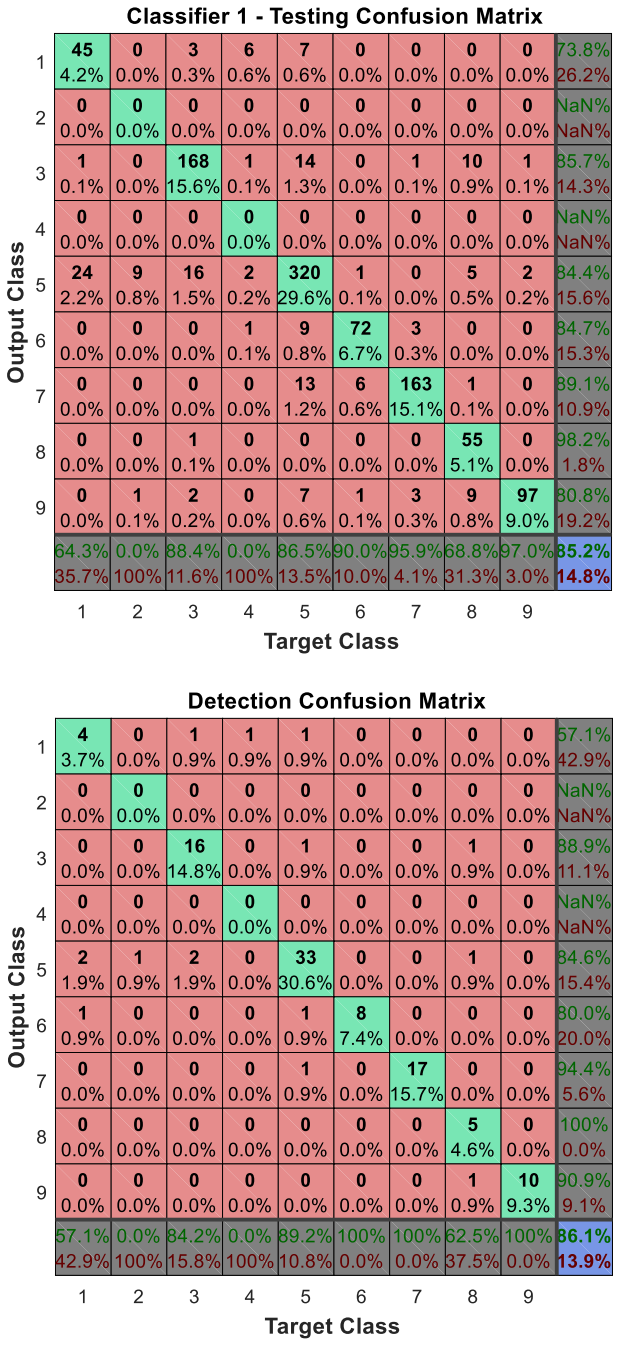
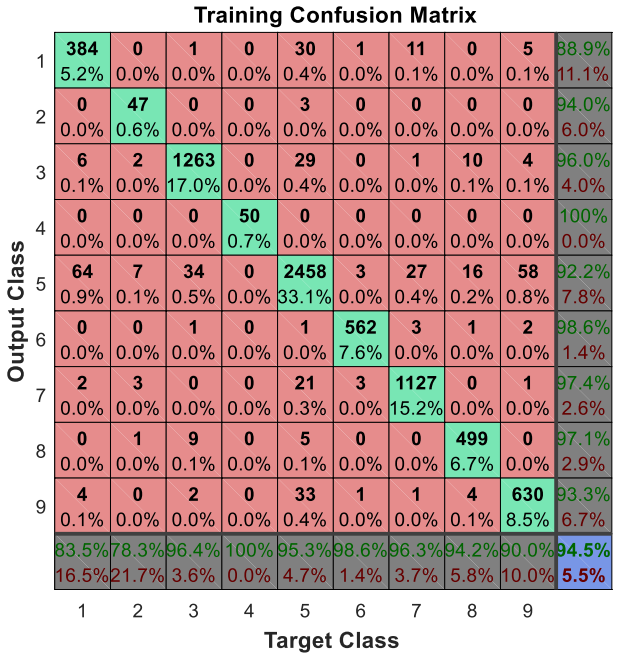
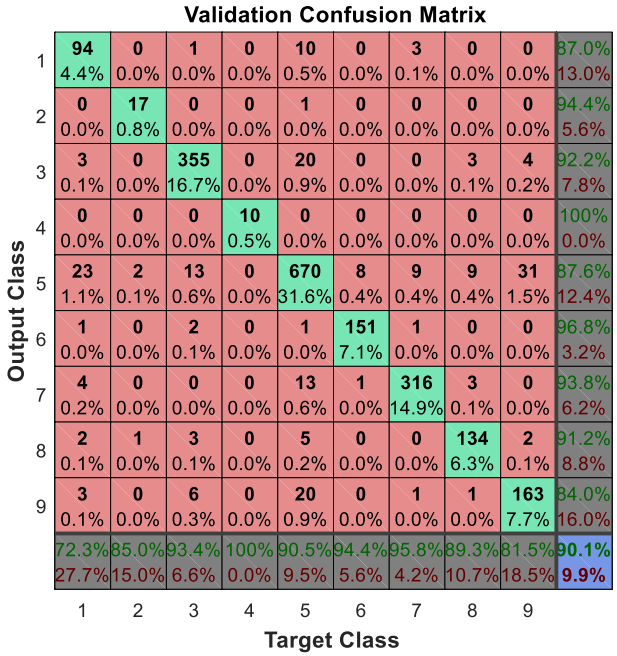


Figure 62: Confusion matrices for Classifier 1

Classifier 2: Wavelet Function: db10





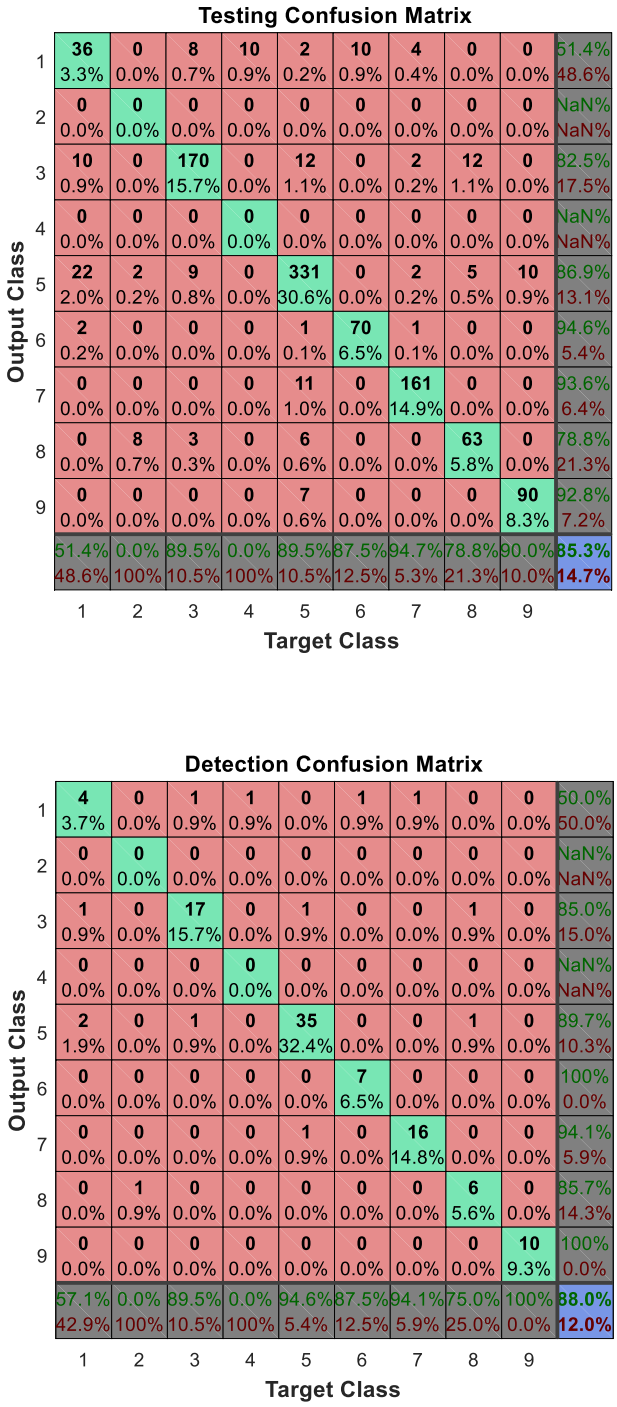
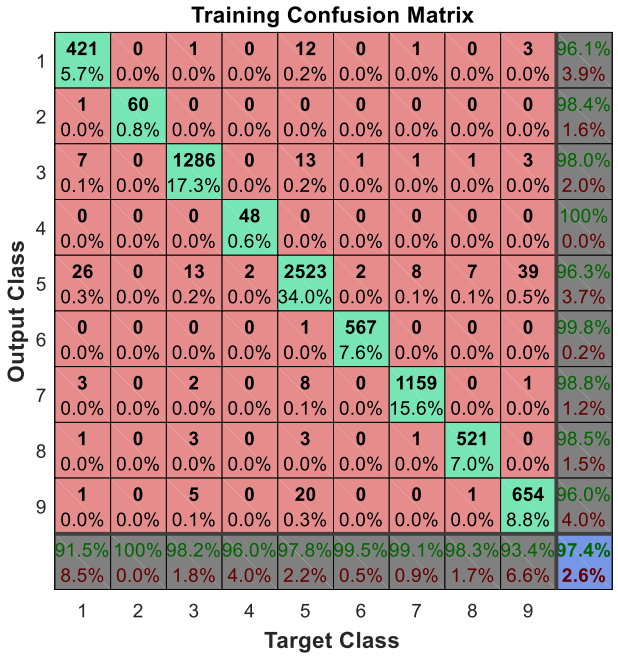
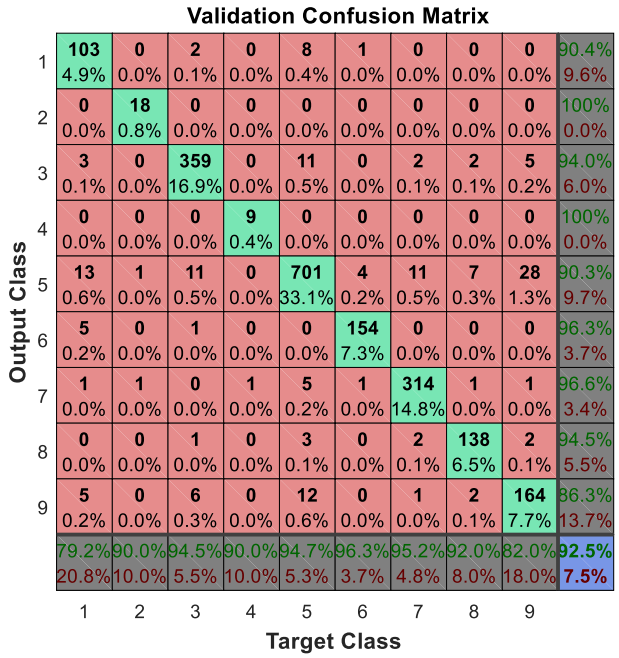
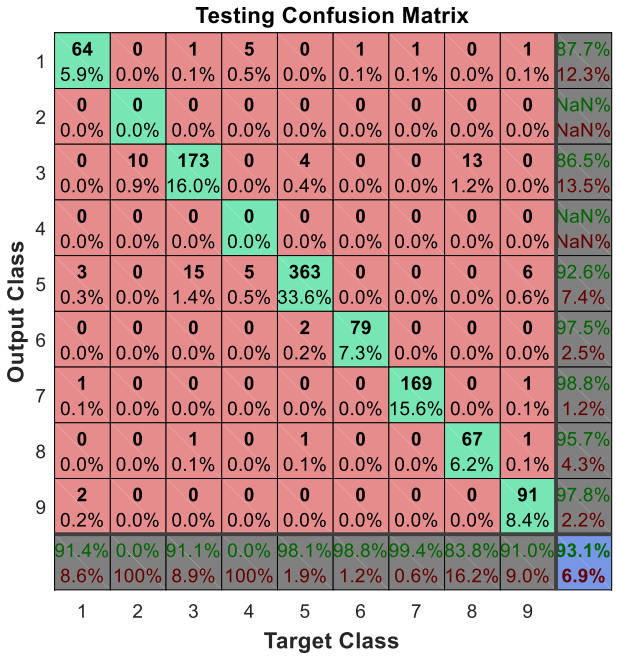


Figure 63: Confusion matrices for Classifier 2

Classifier 3: Wavelet Function: db16







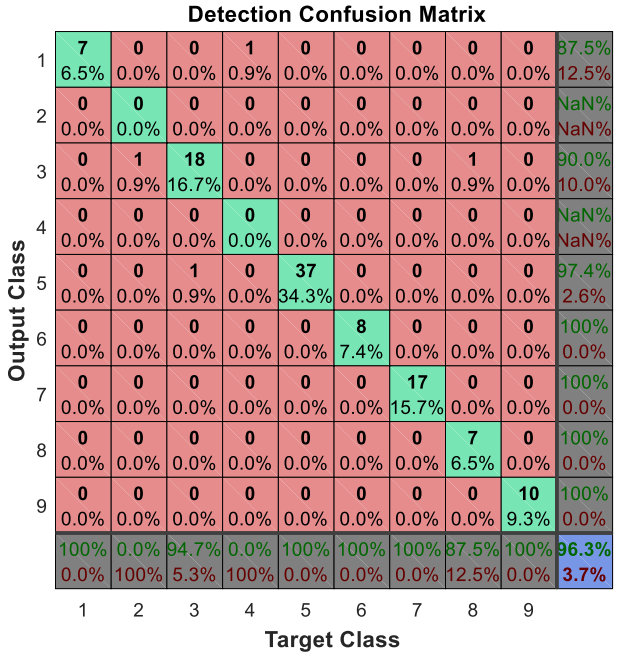


Figure 64: Confusion matrices for Classifier 3

Table 6 presents the performance scores of each classifier for training, validation, testing and detection including precision, recall and F1-sores. The following are the notable observations from the classifier's and their performances.

1. The best overall detection accuracy of 96.3% was achieved by Classifier 3 which was trained on features extracted from db16 wavelet function.
2. All three classifiers were not able to classify any fault representations of fault classes 2 and 4 during testing.
3. Majority voting technique improved the overall accuracy rate for all classifiers which is indicated by the accuracy improvement from testing to detection.

The best detection accuracy of Classifier 3 indicates that the features extracted with db16 wavelet function had the most information that could be inferred to characterize the fault conditions. These results prove that enough information exists in the vibration measurements that can be correlated using in-cylinder pressure measurement to monitor the engine's combustion quality. The performance of Classifiers 1 and 2 indicate that although the information was not as readily available for correlation as it were for Classifier 3, there still exist information that can be utilized for FDD strategy. Even though Classifiers 1 and 2 showed overall accuracy rate that is about 10% lower than that of Classifier 3, more research should be performed to improve their performance rather than ruling them out as potential options for future FDD strategy. The reason is due to the benefits of utilizing db4 or db10 wavelet functions for feature extraction. And the most appealing benefit is the reduction in processing time. As mentioned in Section 5.1, the processing time for

extracting features using db16 is double the amount of time it takes using db4. This could be crucial for when the FDD strategy is to be applied in real-time.

Despite having the high accuracy rates during training and validation across all three classifiers, as shown in Table 6, none of the classifiers were able to detect faults from fault classes 2 and 4. The root cause of this phenomenon is the lack of data for fault classes 2 and 4. The lack of representation of a given fault class in the dataset results in classifier's inability to learn the characteristics that are universal to that specific fault class. The explanation for the very high performance observed during training and very low performance observed during testing could be that the classifiers were able to learn the characteristics of the fault classes but the extent of the engine's variability of behavior was not accounted for due to the limited dataset. Another explanation could be that the classifier was overfitting the dataset that belong to fault classes 2 and 4 due to the imbalance of the dataset. The difference between the two explanations is that the first explanation is reflective of the dataset's inability to provide the classifier the overall picture while the second explanation is reflective of the classifier's neglect of the fault classes 2 and 4. The classifier is trained on the training dataset by updating its weights and biases to reduce the overall accuracy. Therefore, given that fault classes 2 and 4 had very small datasets, they were set as low priority for the classifier to learn as learning the correlation of fault class 5, which has a very high dataset, would yield better accuracy.

The effects of the majority voting technique are reflective in the improvement of overall accuracy observed from testing to detection. As discussed in Section 5.1, a total of

10 baselines were utilized for feature extraction of cycles for both training and testing datasets. The testing confusion matrix accuracy indicates the classifier's ability to detect each feature representation of a single cycle. The majority voting method's function is to mitigate the variations that exist among the 10 feature representations extracted. It can be argued that not all 10 baselines utilized for training ought to be utilized for testing as utilizing fewer baselines for testing would mean shorter processing time of new cycles. However, all 10 baselines were utilized with majority voting method in this research in case the dataset utilized for the development was too small. In the case where the testing accuracy is on par with the accuracy level required to for the FDD strategy the majority voting technique can be omitted.

Table 6: Performance Scores

Classifier 1	Training		**Fault 1**	**Fault 2**	**Fault 3**	**Fault 4**	**Fault 5**	**Fault 6**	**Fault 7**	**Fault 8**	**Fault 9**
		Precision	91.1	100.0	95.4	100.0	91.4	98.9	96.9	97.4	91.4
		Recall	84.6	93.3	95.9	98.0	95.0	97.9	96.5	93.2	86.7
			87.7	96.5	95.7	99.0	93.2	98.4	96.7	95.3	89.0
	Validation		**Fault 1**	**Fault 2**	**Fault 3**	**Fault 4**	**Fault 5**	**Fault 6**	**Fault 7**	**Fault 8**	**Fault 9**
		Precision	85.3	100.0	92.1	100.0	87.4	96.2	94.3	92.4	80.7
		Recall	71.5	80.0	92.3	70.0	91.9	95.6	95.5	88.7	77.5
			77.8	88.9	92.2	82.4	89.6	95.9	94.9	90.5	79.1
	Testing		**Fault 1**	**Fault 2**	**Fault 3**	**Fault 4**	**Fault 5**	**Fault 6**	**Fault 7**	**Fault 8**	**Fault 9**
		Precision	73.8	0.0	85.7	0.0	84.4	84.7	89.1	98.2	80.8
		Recall	64.3	0.0	88.4	0.0	86.5	90.0	95.9	68.8	97.0
			68.7	0.0	87.0	0.0	85.4	87.3	92.4	80.9	88.2
	Detection		**Fault 1**	**Fault 2**	**Fault 3**	**Fault 4**	**Fault 5**	**Fault 6**	**Fault 7**	**Fault 8**	**Fault 9**
		Precision	57.1	0.0	88.9	0.0	84.6	80.0	94.4	100.0	90.9
		Recall	57.1	0.0	84.2	0.0	89.2	100.0	100.0	62.5	100.0
			57.1	0.0	86.5	0.0	86.8	88.9	97.1	76.9	95.2
Classifier 2	Training		**Fault 1**	**Fault 2**	**Fault 3**	**Fault 4**	**Fault 5**	**Fault 6**	**Fault 7**	**Fault 8**	**Fault 9**
		Precision	88.9	94.0	96.0	100.0	92.2	98.6	97.4	97.1	93.3
		Recall	83.5	78.3	96.4	100.0	95.3	98.6	96.3	94.2	90.0
			86.1	85.4	96.2	100.0	93.7	98.6	96.9	95.6	91.6
	Validation		**Fault 1**	**Fault 2**	**Fault 3**	**Fault 4**	**Fault 5**	**Fault 6**	**Fault 7**	**Fault 8**	**Fault 9**
		Precision	87.0	94.4	92.2	100	87.6	96.8	93.8	91.2	84.0
		Recall	72.3	85.0	93.4	100	90.5	94.4	95.8	89.3	81.5
			79.0	89.5	92.8	100.0	89.0	95.6	94.8	90.2	82.7
	Testing		**Fault 1**	**Fault 2**	**Fault 3**	**Fault 4**	**Fault 5**	**Fault 6**	**Fault 7**	**Fault 8**	**Fault 9**
		Precision	51.4	0.0	82.5	0.0	86.9	94.6	93.6	78.8	92.8
		Recall	51.4	0.0	89.5	0.0	89.5	87.5	94.7	78.8	90.0
			51.4	0.0	85.9	0.0	88.2	90.9	94.2	78.8	91.4
	Detection		**Fault 1**	**Fault 2**	**Fault 3**	**Fault 4**	**Fault 5**	**Fault 6**	**Fault 7**	**Fault 8**	**Fault 9**
		Precision	50.0	0.0	85.0	0.0	89.7	100.0	94.1	85.7	100
		Recall	57.1	0.0	89.5	0.0	94.6	87.5	94.1	75.0	100
			53.3	0.0	87.2	0.0	92.1	93.3	94.1	80.0	100.0
.Classifier 3	Training		**Fault 1**	**Fault 2**	**Fault 3**	**Fault 4**	**Fault 5**	**Fault 6**	**Fault 7**	**Fault 8**	**Fault 9**
		Precision	96.1	98.4	98.0	100.0	96.3	99.8	98.8	98.5	96.0
		Recall	91.5	100.0	98.2	96.0	97.8	99.5	99.1	98.3	93.4
			93.7	99.2	98.1	98.0	97.0	99.7	99.0	98.4	94.7
	Validation		**Fault 1**	**Fault 2**	**Fault 3**	**Fault 4**	**Fault 5**	**Fault 6**	**Fault 7**	**Fault 8**	**Fault 9**
		Precision	90.4	100.0	94.0	100.0	90.3	96.3	96.6	94.5	86.3
		Recall	79.2	90.0	94.5	90.0	94.7	96.3	95.2	92.0	82.0
			84.4	94.7	94.3	94.7	92.5	96.3	95.9	93.2	84.1
	Testing		**Fault 1**	**Fault 2**	**Fault 3**	**Fault 4**	**Fault 5**	**Fault 6**	**Fault 7**	**Fault 8**	**Fault 9**
		Precision	87.7	0.0	86.5	0.0	92.6	97.5	98.8	95.7	97.8
		Recall	91.4	0.0	91.1	0.0	98.1	98.8	99.4	83.8	91.0
			89.51	0.0	88.74	0.0	95.27	98.15	99.10	89.36	94.28
	Detection		**Fault 1**	**Fault 2**	**Fault 3**	**Fault 4**	**Fault 5**	**Fault 6**	**Fault 7**	**Fault 8**	**Fault 9**
		Precision	87.5	0.0	90.0	0.0	97.4	100.0	100.0	100.0	100.0
		Recall	100.0	0.0	94.7	0.0	100.0	100.0	100.0	87.5	100.0
			93.3	0.0	92.3	0.0	98.7	100.0	100.0	93.3	100.0

Chapter 6: Conclusion

The purpose of this research was to develop a FDD strategy that can monitor a combustion related fault of an internal combustion engine on a cycle-by-cycle basis. This research was performed in collaboration with Ford Powertrain Engineering Research and Development Center as a continuation of the ongoing FDD research conducted at CMHT. This section will summarize the research, highlight the notable contributions, and provide suggestions for future work.

6.1 Research Contributions

In this research, a FDD strategy consisting of data acquisition, data pre-processing, feature extraction and classifier development specific sub strategies was proposed. The dataset required for the development of the FDD strategy was proposed to be acquired from currently existing testing procedures and testing facilities. During this research, the dataset consisting of information rich sensor measurements were obtained. These measurements function to provide insight into what is occurring inside the engine. In this research, the combustion quality was the characteristic being monitored by the FDD strategy. Therefore, the state of the engine's combustion quality was obtained through the analysis of the in-cylinder pressure measurement. The combustion quality indicated by analysis of the pressure measurement functioned as the ground truth. However, due to the invasive and

expensive nature of the pressure transducers, they are only available in the testing environment. Therefore, a vibration measurement was proposed to be the monitored measurement for the FDD strategy due to the non-invasive and cheaper cost of accelerometers. The difficulty in inferring the combustion quality using vibration measurement was proposed to be mitigated during the classifier development through labeling of vibration measurement using information rich pressure measurement. To help with the dimensionality reduction and correlation of vibration data to the combustion quality, feature extraction using Extended-MSPCA was performed on vibration measurements in crank angle domain. Given the features extracted and its corresponding combustion quality obtained through pressure measurement analysis, the correlation between the features and the combustion quality was developed using the machine learning algorithm through supervised learning.

In summary, the following are the major contributions of this research:

1. Development of new data acquisition procedures for combustion related faults that seamlessly integrates into engine testing processes to obtain high quality engine data.
2. Development of data pre-processing techniques unique to internal combustion engines that allows analysis in crank angle domain on a cycle by cycle basis.
3. Development of vibration measurement labeling procedure through analysis of in-cylinder pressure measurement.

4. Application of Extended-MSPCA on vibration measurement using multiple baseline techniques customized for internal combustion engines to mitigate variable nature of the dataset.
5. Development a of classifier using ANN and majority voting technique that has shown to be able to classify new dataset for combustion quality with 96.3% accuracy.

This research demonstrated that there are creative ways to utilized the data that are being collected during testing processes to enable advancements in technology through data analysis. More specifically, the research has shown that complex relationship between pressure and vibration of an engine can be extracted to develop a FDD strategy that is capable of monitoring the engine's combustion quality on a cycle by cycle basis.

6.2 Future Work

This research was conducted to assess the feasibility of the proposed FDD strategy in monitoring the engine's combustion quality. The development of the proposed FDD strategy was focused more on the successful application rather than the fine tuning of parameters in each of the sub-strategies. Therefore, there are many aspects of the proposed FDD strategy that can be researched to understand the capabilities of the strategy itself and how it can be fine-tuned to improve the its performance or to be applied to different types of faults.

Given the set of data that was collected in this research, the following are the recommendations for future research:

1. Study of different wavelet families on the existing dataset for feature extraction.
2. Study of different pressure analysis methods for data labeling of combustion related fault conditions.
3. Study of different baseline selection methods and determination of the optimal number of baselines for feature extraction.
4. Development of data imbalance mitigation methods.
5. Development of data processing procedures for continuous monitoring in real time applications.

In order to bring this technology to life, a continuous research approach is recommended where a set of data collection protocols are developed to acquire coherent data sets from different types of engines. This protocol is envisioned to standardize data collection properties such as the types of measurements being collected and the preferred location of the transducers. Data collection protocols may differ for different types of faults being monitored. In addition, the protocol should consist of generalized rules to enable application to different types of engines. With data from different types of engines, a research can be conducted to see if a FDD strategy developed for one engine can be applied to a different engine or different type of engine.

References

[1] R. Isermann, Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance, Springer, 2006.

[2] P. K. K. S. K. Yadav, "Condition Monitoring of Internal Combustion Engine Using EMD and HMM," in *Intelligent Autonomous Systems*, Springer, 2010, pp. 167-185.

[3] R. Stone, Introduction to Internal Combustion Engines, London: Palgrave, 1999.

[4] J. Heywood, Internal Combustion Engine Fundamentals, McGraw-Hill Education, 1988.

[5] S. Fallah, Electric and Hybrid Vehicles - Technologies, Modeling and Control: A Mechatronic Approach, Wiley-Blackwell, 2014.

[6] Nuclear Power contributors, "Otto Cycle – Otto Engine," Nuclear Power , [Online]. Available: https://www.nuclear-power.net/nuclear-engineering/thermodynamics/thermodynamic-cycles/otto-cycle-otto-engine/. [Accessed 2019].

[7] J. A. Caton, "The Thermodynamics of Internal Combustion Engines: Examples of Insights," *Inventions*, vol. 3, no. 2, p. 33, 2018.

[8] N. Hall, "Ideal Otto Cycle," National Aeronautics and Space Administration, 05 May 2015. [Online]. Available: https://www.grc.nasa.gov/www/k-12/airplane/otto.html. [Accessed 2019].

[9] S. K. Y. a. P. K. Kalra, "Condition Monitoring of Internal Combustion Engine Using EMD and HMM," in *Intelligent Autonomous Systems*, Springer, 2010, pp. 167-185.

[10] V. Venkatasubramanian, "A review of process fault detection and diagnosis Part III: Process history based methods," *Elsevier - Computers & Chemical Engineering*, pp. 327-346, 2003.

[11] V. Venkatasubramanian, "A Review of Process Fault Detection and Diagnosis Part 1: Quantitative Model-Based Methods," *Elsevier - Computers & Chemical Engineering* , 2003.

[12] M. Ismail, "Industrial Extended Multi-Scale Principle Components Analysis for Fault Detection and Diagnosis of Car Alternators and Starters," McMaster University , 2015.

[13] J. Tjong, "Engine dynamic signal monitoring and diagnostics," *University of Windsor,* 1993.

[14] H. W. Miller, "Piston slap detection in combustion engines using unique signal analysis techniques," *University of Windsor,* 1989.

[15] S. Delvecchio, "Condition Monitoring in Diesel Engines for Cold Test Applications. Part I: Vibration Analysis for Pass/Fail Decision," 2007.

[16] R. Tailony, "Internal Combustion Engine Cold Testing: Modeling, Analysis, and Development," University of Toledo, 2019.

[17] L. Wei, "Modeling and Simulation of System Dynamics for Cold Test," *Second Inernational Symposium on Intelligent Information Technology Application,* 2008.

[18] B. Ray, "Engine defect defection using wavelet analysis," University of Windsor , 2007.

[19] E. Leitzinger, "Development of in-process engine defect detection methods using NVH indicators," University of Windsor , 2002.

[20] I. Arasaratnam and S. Habibi, "Engine Fault Detection Using Vibration Signal Reconstruction in the Crank Angle Domain," *SAE International,* 2011.

[21] ATS Automation , "Powertrain Final Test," ATS Automation, [Online]. Available: https://www.atsautomation.com/en/Transportation/Testing%20Automation/Powertrain%20Final%20Test.aspx#cold. [Accessed December 2019].

[22] A. Doghri, "A Real Time Fault Detection and Diagnosis System for Automotive Applications," *McMaster University*, 2019.

[23] S. Hodgins, "A Wireless Sensor for Fault Detection and Diagnosis of Internal Combustion Engines," McMaster University, 2017.

[24] S. R. Haqshenas, "Multiresolution - multivariate analysis of vibration signals: application in fault diaagnosis of internal combustion engines," McMaster University, 2012.

[25] N. Gunnemann, "Predicting Defective Engines using Convolutional Nueral Networks on Temporal Vibration Signals," *Proceedings of Machine Learning Research*, vol. 74, p. 92:102, 2017.

[26] Y. Shantnawi, "Fault diagnosis in Internal Combustion Engines Using Extension Neural Network," *IEEE Transactions on Industrial Electronics*, pp. 1434- 1443, 2014.

[27] A. S. G. Chandroth, "Vibration Signatures, Wavelets and Principal Components Analysis in Diesel Engine Diagnostics," *WIT Transactions on The Built Environment*, vol. 45, p. 10, 1999.

[28] Q. Gao, C. Fan and Q. Meng, "Rotating machine fault diagnosis using empirical mode decomposition," *Mechanical System and Signal Processing*, no. 22, pp. 1072-1081, 2008.

[29] J. Wu and C. Liu, "Investigation of engine fault diagnosis using discrete wavelet transform and neural network," *Expert Systems with Applications*, vol. 35, pp. 1200-1213, 2008.

[30] A. Sharkey, O. Gopinath and Chandroth , "Acoustic emission, cylinder pressure and vibration: a multisensor approach to robust fault diagnosis," *University of Sheffield*, 2000.

[31] D. Mba and R. Rao, "Development of Acoustic Emission Technology for Condition Monitoring and Diagnosis of Rotating Machines; Bearings, Pumps, Gearboxes, Engines and Rotating Structures," *The Shock and Vibration Digest* , vol. 38, pp. 3-16, 2006.

[32] J. Ben-Ari, "Vibration Signature Analysis as a Fault Detection Method for SI Engines," *SAE* , 1993.

[33] R. B. W. Heng, "Statistical Anlysis of Sound and Vibration Signals for Monitoring Rolling Element Bearing Condition," *Elsevier: Applied Acoustics,* vol. 53, pp. 211-226, 1998.

[34] S. Zhang, Machine Learning and Deep Learning Algorithms for Bearing Fault Diagnostics - A Comprehensive Review, 2019.

[35] W. Zhang, "A New Deep Learning Model for Fault Diagnosis with Good Anti-Noise and Domain Adaptation Ability on Raw Vibration Signals," *MDPI,* 2017.

[36] M. D. Prieto, "Bearing Fault Detection by a Novel Condition-Monitoring Scheme Based on Statistical Time Features and Neural Networks," *IEE Transactions on Industrial Electronics,* vol. 30, no. 8, 2013.

[37] S. Hunt, "Engine Vibration of a Compressed Natural Gas Engine as an effect of Combustion," *University of Windsor,* 2016.

[38] I. Arasaratnam, "Engine Fault Detection Using Vibration Signal Reconstruction in the Crank-Angle Domain," *SAE International,* 2011.

[39] S. E. Pandarakone, "A Comparative Study between Machine Learning Algorithm and Artificial Intelligence Neural Network in Detecting Minor Bearing Fault of Induction Motors," *MDPI,* 2019.

[40] I. Amihai, "An Industrial Case Study Using Vibration Data and Machine Learning to Predict Asset Health," *IEEE,* 2018.

[41] M. A. Lilo, "Identify and Classify Vibration Fault Based on Artificial Intelligence Techniques," *Journal of Theoretical and Applied Information Technology,* vol. 94, no. 2, 2016.

[42] S. Seidlitz, "An Optimization Approach to Valvetrain Design," *SAE Technical Paper,* 1990.

[43] N. Politis, "Advanced Time-Frequency Analysis Applications in Earthquake Engineering," 2003.

[44] Muller-BBM VibroAkustik System, *Rotational Analysis in the Angular Domain,* Planegg: Muller-BBM VibroAkustik System, 2016.

[45] National Instruments, "Getting Started with Order Analysis," National Instruments Corporation, 2007.

[46] G. Sosa, "Automatic detection of wheezes by evaluation of multiple acoustic feature extraction methods and C-weighted SVM," *Progress in Biomedical Optics and Imaging - Proceedings of SPIE,* vol. 9287, 2014.

[47] R. Polikar, "The Engineers Ultimate guide to Wavelet Analysis: The Wavelet Tutorial," 11 5 2006. [Online]. Available: https://cseweb.ucsd.edu/~baden/Doc/wavelets/polikar_wavelets.pdf. [Accessed 2019].

[48] I. Daubechies, "Where do Wavelets Come From?," *IEEE*, vol. 84, no. 4, pp. 510-513, 1996.

[49] V. Dhandapani and R. Seshasayanan, "Power-optimized log-based image processing system," *EURASIP Journal on Image and Video Processing* , p. 37, 2014.

[50] Hackernoon, "Principal Component Analysis - Unsupervised Learning Model," Hackernoon, Jan 2019. [Online]. Available: https://hackernoon.com/principal-component-analysis-unsupervised-learning-model-8f18c7683262. [Accessed Dec 2019].

[51] T. Villegas, M. Fuente and M. Rodriguez, "Principal Component Analysis for Fault Detection and Diagnosis. Expereience with a pilot plant," *Advances in Computerational Intelligence, Mac-Machine Systems and Cybernetics* , pp. 147-152, 2010.

[52] G. A. Cherry and S. J. Qin, "Monitoring Non-normal Data with Principal Component Analysis and Adaptive Density Estimation," *IEEE Conference on Decision and Control*, pp. 352-359, 2007.

[53] J. Jackson, in *A User's guide to principal components*, New York, John Wiley and Sons Inc., 1991.

[54] G. Box, "Some theorems on quadratic forms applied in the study of analysis of variance problems, i. effect of inequality of variance in the one-way classification," *The Annals of Mathematical Statistics* , vol. 2, no. 25, pp. 290-302, 1954.

[55] B. Bakshi, "Multiscale PCA with application to multivariate statistical process monitoring," *AlChE,* vol. 44, no. 7, 2004.

[56] S. Yoon and M. J, "Principal-Component Analysi sof Multiscale Data for Process Monitoring and Fault Diagnosi," *AlChE Journal,* vol. 50, no. 11, pp. 2891-2903, 2004.

[57] S. Haykin, Neural Networks and Learning Machines, Hamilton: Pearson, 1999.

[58] Z. Benedek, "Pushrod vs. Overhead cam engines," 19 Jan 2018. [Online]. Available: https://allaboutdieselz.blogspot.com/2018/01/pushrod-vs-overhead-cam-engines.html. [Accessed Dec 2019].

[59] E. Webb, "Introduction to Internal Combustion Engines," 2019. [Online]. Available: https://slideplayer.com/slide/13027058/. [Accessed Dec 2019].

[60] R. Muscoplat, "4.6L V8 Ford firing order," Rick's Auto Repair Advice, 2009. [Online]. Available: https://ricksfreeautorepairadvice.com/4-6l-v8-ford-firing-order/. [Accessed 2019].

[61] National Instruments, "Encoder Measurements: How-To Guide," National Instruments, 9 Oct 2019. [Online]. Available: http://www.ni.com/tutorial/7109/en/. [Accessed 2019].

[62] S. B. Quang, "Rotary Encoder Training Material," Maruei Vietnam Precision, 2013.

[63] AVL, "Crank Angle Encoder 365-Series Prodcut Description," AVL, Feb 2010. [Online]. Available: https://www.avl.com/documents/10138/885965/angle-encoder-365C-PS-2010-ENG+neu.pdf. [Accessed 2019].

[64] Kistler, "Test & Measurement Pressure," Kistler, 2018. [Online]. Available: https://www.kistler.com/?type=669&fid=541&model=download&callee=frontend. [Accessed 2019].

[65] AVNET ABACUS, "Piezoelectric pressure sensors," AVNET ABACUS, [Online]. Available:

https://www.avnet.com/wps/portal/abacus/solutions/technologies/sensors/pressure-sensors/core-technologies/piezoelectric/. [Accessed Dec 2019].

[66] Kistler, "PiezoStar Pressure Sensor," Kistler, 2018. [Online]. Available: https://www.kistler.com/?type=669&fid=102821. [Accessed Dec 2019].

[67] National Instruments, "Measuring Vibration with Accelerometers," National Instruments, 14 Mar 2019. [Online]. Available: https://www.ni.com/en-ca/innovations/white-papers/06/measuring-vibration-with-accelerometers.html. [Accessed Dec 2019].

[68] Dytran, "Performance Specification Triaxial IEPE Acceleromter Model Number 3313A1," 2010.

[69] AutoZone, "Knock Sensor Repair Guide," AutoZone, [Online]. Available: https://www.autozone.com/repairguides/Silverado-2008/Components-Systems/Knock-Sensor-KS/_/P-0996b43f81b3c351. [Accessed 2019].

[70] I. Dibrani, "The new V8 POwer Unit for the BMW M3," BMW Heaven, 8 April 2007. [Online]. Available: http://www.bmwheaven.com/index2.php?option=com_content&task=view&id=119&pop=1&page=6&Itemid=3. [Accessed 2019].

[71] National Instruments, "Introduction to Data Acquisition," National Instruments, 14 Mar 2019. [Online]. Available: https://www.ni.com/en-ca/innovations/white-papers/06/introduction-to-data-acquisition.html#section-1450863631. [Accessed Dec 2019].

[72] National Instruments, "How to Choose the Right DAQ Hardware for Your Measurement System," National Instruments, 5 Mar 2019. [Online]. Available: https://www.ni.com/en-ca/innovations/white-papers/11/how-to-choose-the-right-daq-hardware-for-your-measurement-system.html. [Accessed Dec 2019].

[73] W. C. Chang, "An Improved Method of Investigation of Combustion Parameters in a Natural Gas Fuelled SI Engine with EGR and H2 as Additives," University of Birmingham, 2002.

[74] B. T. Narendiranath, "Journal Bearing Fault Detection Based on Daubechies Wavelet," *Archives of Acoustics,* vol. 42, no. 3, pp. 401-414, 2017.

[75] C. S. Burrus, "Fast Approximate Fourier Transform via Wavelets Transform," *Proceedings of SPIE - The International Society for Optical Engineering* , 1997.

[76] W. Ribbens, "Chapter 4 - The Basics of Electronic Engine Control," *Understanding Automotive Electronics* , pp. 135-182, 2017.

[77] P. Niedenfuhr, "The Piezoresistive Effect and Measuring Pressure," FirstSensor, Feb 2019. [Online]. Available: https://blog.first-sensor.com/en/piezoresistive-effect. [Accessed Dec 2019].

[78] E. Walter and R. Walter, "On-Board Diagnostics Background and Standards," in *OBD and CAN data acquisition for light-duty vehicles*, SAE, 2018, pp. 25-36.

[79] AutoTap, "OBDII: Past, Present & Future," AutoTap, [Online]. Available: http://www.autotap.com/techlibrary/obdii_past_present_future.asp. [Accessed DEc 2019].

[80] J. Faiz, "A Survey on Condition Monitoring and Fault Diagnosis in Line-start and Inverter-fed Broken Bar Induction Motors," *IEEE International Conference on Power Electronics, Drives and Energy Systems* , 2012.

[81] B. Yazici, "An Adaptive Statistical Time-Frequency Method for Detection of Broken Bars and Beraing using the Stator Current," *IEEE Transaction on Industry Applications,* vol. 35, no. 2, pp. 442-452, 199.

[82] G. Yen, "Conditional Health Monitoring using Vibration Signatures," *IEEE Proceedings of the 38th Conference on Decision and Control* , pp. 4493-4498, 1999.

[83] S. Goumas, "Classification of Washing Machine Vibration Signals Using Discrete Wavelet Analysis for Feature Extraction," *IEEE Transaction on Instrumentation and Measurements,* vol. 51, no. 3, pp. 497-508, 2002.

[84] I. Djurovic, "A Virtual Instrument for Time-Frequency Analysis," *IEEE Transcation on Instrumentation and Measurements,* vol. 48, no. 6, pp. 1086-1092, 1999.

[85] A. Chen, "Internal combustion engine vibration analysis with short-term Fourier-transform," *IEEE 2010 3rd International Congress on Image and Signal Processing*, 2010.