Big Data Analysis Report
P002: IoT-based Virtual Health Determination

Department: Department of Statistics

Student ID:

郭依璇 (410978004) (Yi-Hsuan Kuo)

戴衣伶 (410978038)

林瑋珈 (410978051)

薛珮妤 (410978055)

李博業 (410978058) (Bo-Ye, Li)

侯言蓁 (411078018)

陳皓鈞 (411078038)

Advisor: Ping-Yang Chen, Ph.D.

June, 2024

Table of Contents

1 Introduction

In modern manufacturing, robotic arms play a crucial role; however, any issues in their transmission systems can directly affect production efficiency and product quality. Current inspection procedures require downtime and significant human labor, and they are limited in preventing sudden failures. Therefore, this project aims to establish a robotic arm health monitoring system based on vibration sensor data to improve production efficiency and reduce losses caused by equipment malfunctions. By employing machine learning methods, we will investigate the relationship between vibration signals and transmission health, in order to develop a predictive maintenance mechanism and create a more efficient and reliable production environment for the manufacturing industry.

2 Project Background and Purposes

2.1 Project Background

Robotic arms are common automation devices in industrial production, typically performing critical tasks such as assembly, material handling, and welding. However, due to prolonged operation and wear, robotic arms may encounter various issues, including bearing wear, looseness, and imbalance, which can lead to reduced performance, malfunctions, or even damage.

Vibration sensors can be used to monitor the vibration behavior of robotic arms. By analyzing vibration data, it is possible to assess the operating conditions and overall health of the robotic arm. Vibration data contains rich information, such as frequency, amplitude, and waveform, which can be used to detect abnormal vibrations and evaluate the health status of the equipment.

2.2 Project Purposes

By performing real-time monitoring and analysis of vibration sensor data, it is possible to achieve timely diagnosis and prediction of the health status of robotic arms. When abnormal vibrations are detected, the system can issue alerts and prompt maintenance actions, thereby reducing the risk of equipment damage and production interruptions, and improving both equipment reliability and production efficiency.

3 Project Implementation Plan

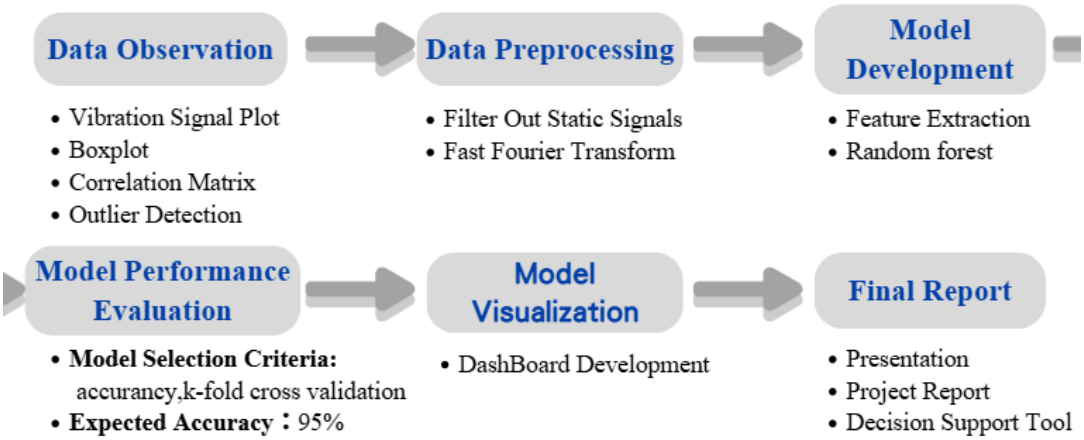

Fig 1: Execution Plan

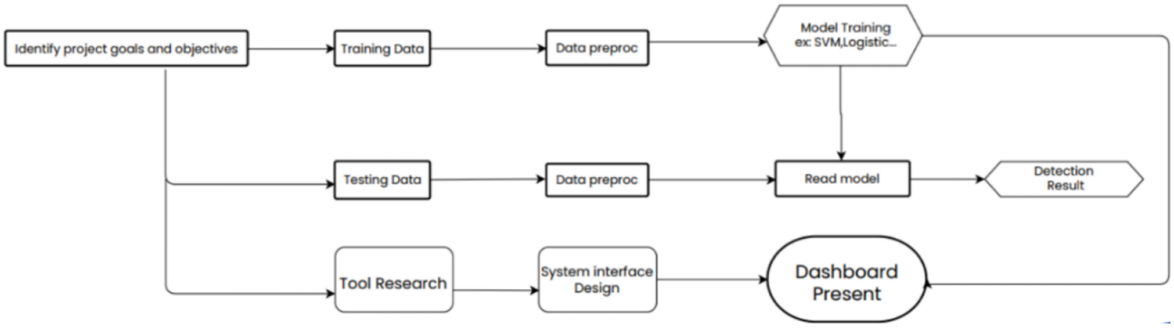

Fig 2: Flowchart

Week / Task Name	Second Semester of Academic Year 113													
	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Literature Review														
Data Observation and Preprocessing														
Model Design Concept														
First Course Report														
Model Testing and Analysis Methods														
Model Performance Verification and Adjustment														
Second Course Report														
Verification Result														
Model Results Visualization														
Third Course Report														
Project Report														

Fig 3: Gantt Chart

4 Data Exploration

4.1 Data Description

The data were collected from four tri-axial accelerometers installed on the robotic arm, capturing vibration signals along the x, y, and z axes. The accelerometers were installed as follows:

- **Xa** – Motor side of the horizontal-axis transmission of the robotic arm
- **Xb** – Idler side of the horizontal-axis transmission of the robotic arm
- **Ya** – Motor side of the vertical-axis transmission of the robotic arm
- **Yb** – Idler side of the vertical-axis transmission of the robotic arm

Different levels of load were applied to the transmission shafts in different directions as follows:

- **Horizontal direction:** 65, 80, 95, 130, where the normal load is approximately 80
- **Vertical direction:** 220, 260, 300, 380, where the normal load is approximately 260

Data acquisition setup:

Under a fixed tension, the robotic arm operated at a constant speed and mode for approximately 5–6 minutes. Vibration signals were recorded every 5 seconds, with each recording saved as a separate file.

4.2 Data Exploration

(1) Vibration Signal Plot – Outliers

The vibration signals recorded by the accelerometers on the robotic arm were plotted as vibration signal graphs (see Figure 4) to observe the characteristics of the data. It was found that the vibration amplitude of the robotic arm varies depending on whether the arm is in motion or stationary.

When the robotic arm is temporarily stationary, the vibration signal exhibits a stable state, approximately ranging from 0.08 to 3. This variation is due to differences in acceleration amplitudes (g) caused by the installation positions of the accelerometers. However, during operation, the acceleration amplitude forms peaks that are about 2 to 3 times larger than the baseline non-vibration signal.

Therefore, before analysis, it is necessary to remove the stable (stationary) data. Only the vibration data recorded during the operation of the robotic arm are used for analysis to prevent the model from being affected by a large amount of non-

operational data.

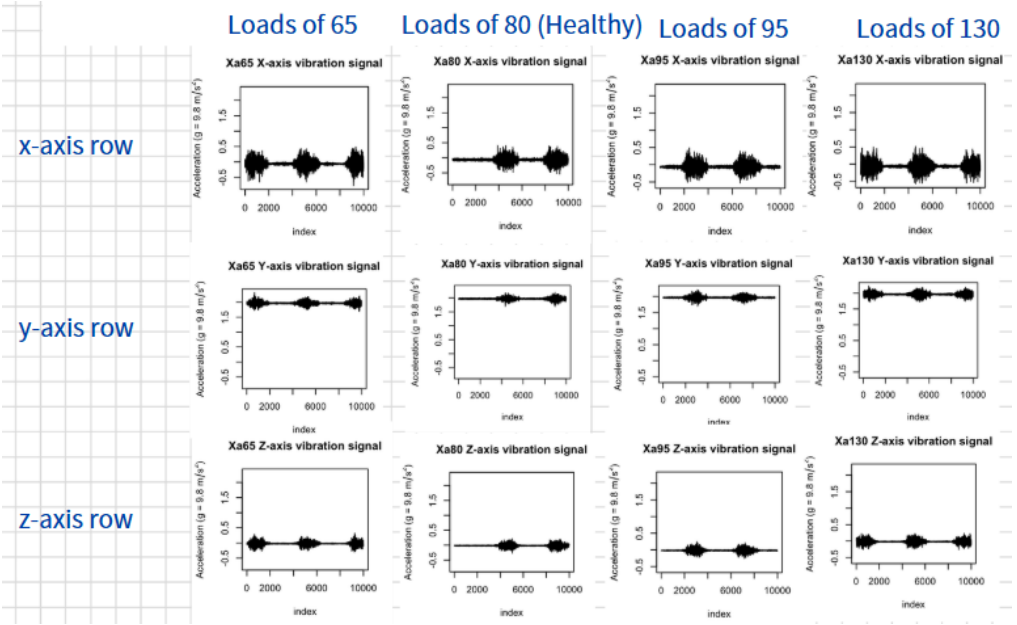

Fig 4: vibration signal graphs

(2) Boxplot

From the boxplots, it can be observed that the tri-axial vibrations measured by accelerometers installed at different positions operate on the same baseline under various loads, making it difficult to distinguish differences by visual inspection alone. Only slight variations in vibration amplitude can be noted.

For example, for the accelerometer installed on the horizontal-axis motor side (Xa), the X-axis vibration under a load of 65 exhibits lower acceleration compared to the X-axis vibration under 80 (healthy state), while the Z-axis shows higher acceleration. On the idler side (Xb), the X-axis vibration under a load of 95 shows a larger range of acceleration.

Similarly, for the vertical-axis operation, the accelerometer on the motor side (Ya) under a load of 220 shows lower X-axis acceleration compared to the healthy load of 260. The Y-axis vibration under loads 220 and 300 has a smaller acceleration range than under 260, while the Z-axis vibrations under all three non-healthy loads exhibit a shorter acceleration range (see Figure 5). On the idler side (Yb), under a load of 220, the X-axis acceleration is lower compared to the healthy load of 260, whereas the Y- and Z-axis vibrations under loads 300 and 380 show a larger acceleration range than under 260 (see Figure 6).

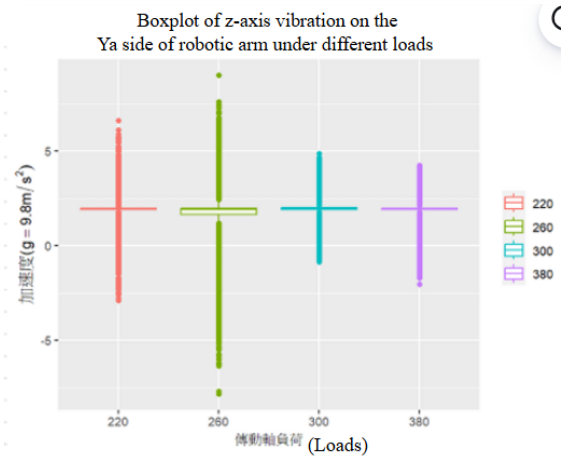

Fig 5

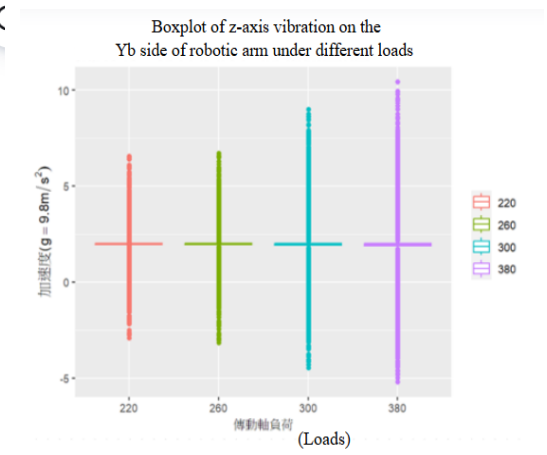

Fig 6

(3) Correlation Matrix

When examining the four accelerometers (Xa, Xb, Ya, and Yb) under all load conditions, the correlations among the X-, Y-, and Z-axes show no substantial variation.

For the accelerometer installed at Xa, a weak positive correlation is observed between the X- and Z-axes. In contrast, the three axes at Xb exhibit near-zero correlation (as shown in Fig. 7).

Under vertical operation, the accelerometers installed at Ya and Yb display slight positive correlations between the X- and Y-axes (as shown in Fig. 8).

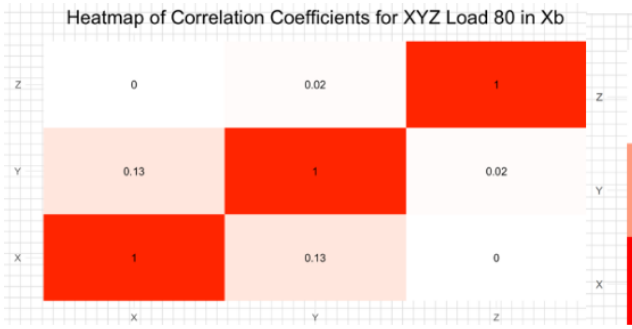

Fig 7

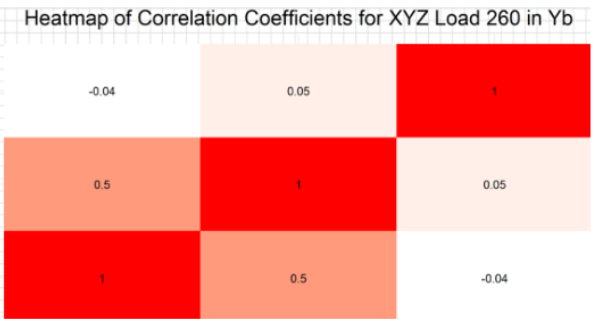

Fig 8

(4) Data Visualization

During the exploratory data analysis, we observed that the waveforms under vertical operation exhibit more pronounced differences across varying load conditions compared to those under horizontal operation.

For instance, at the motor side, the waveform under a load of 260 appears denser than that under a load of 220 (as shown in Fig. 9). Similarly, at the idler side, the waveform under a load of 380 is denser than that under a load of 220 (as shown in Fig. 10).

Based on these observations, we initially hypothesized that vertical operation might yield better classification performance than horizontal operation.

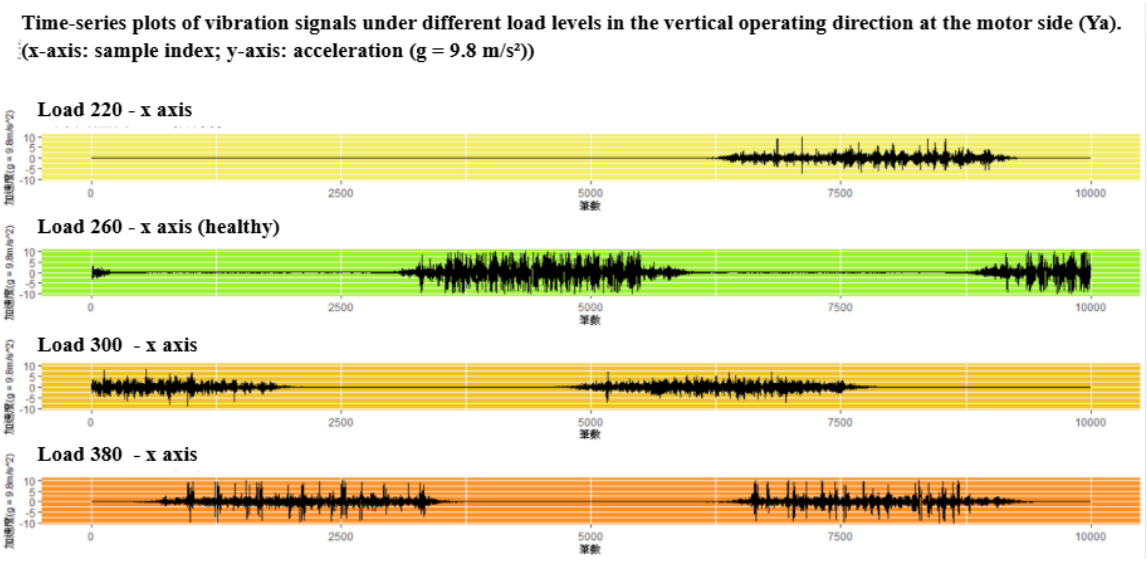

Fig 9: Time-Series Plot of Ya Data

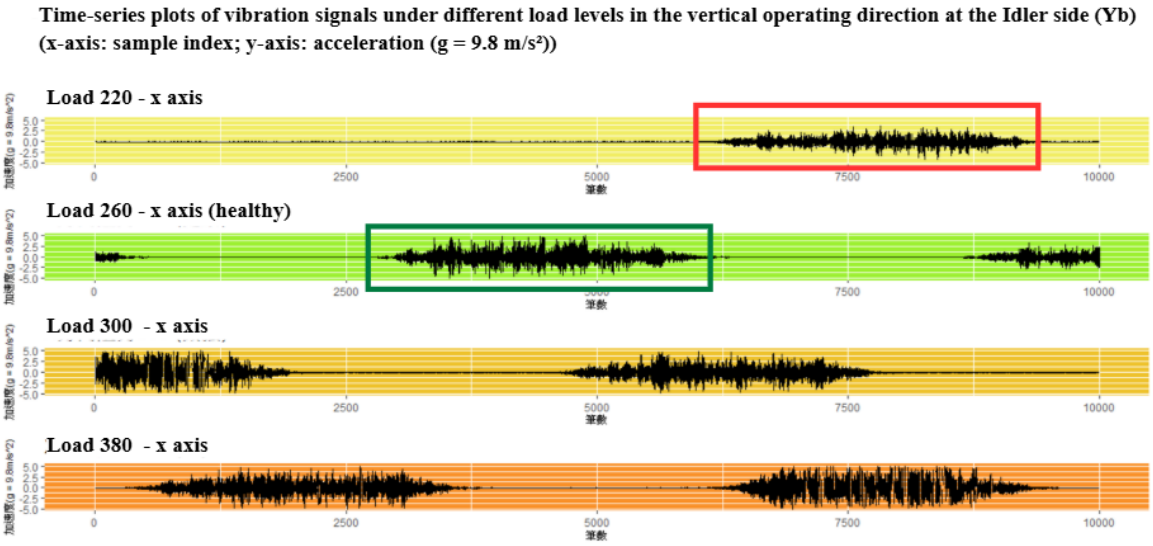

Fig 10: Time-Series Plot of Yb Data

5 Data Preprocessing

Through a review of the literature, we found that robotic arms gradually undergo wear and degradation over time, which leads to deviations in the vibration signals they generate, a phenomenon referred to as **degradation**. In addition, as noted in the data exploration stage, the vibration data recorded during non-operational (idle) periods of the robotic arm fall outside the scope of this study.

To reduce potential errors in subsequent diagnostic results, the preprocessing stage therefore aims to verify the presence of degradation and to remove steady-state (idle) data from the dataset.

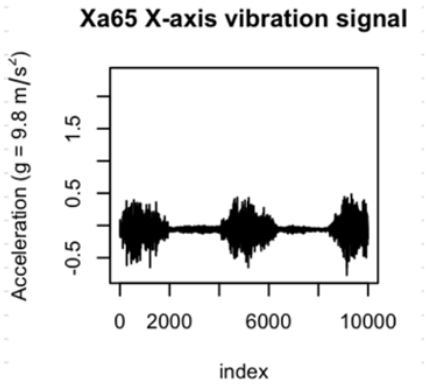

Fig 11: Schematic Diagram of the Raw Data

5.1 Check for Degradation

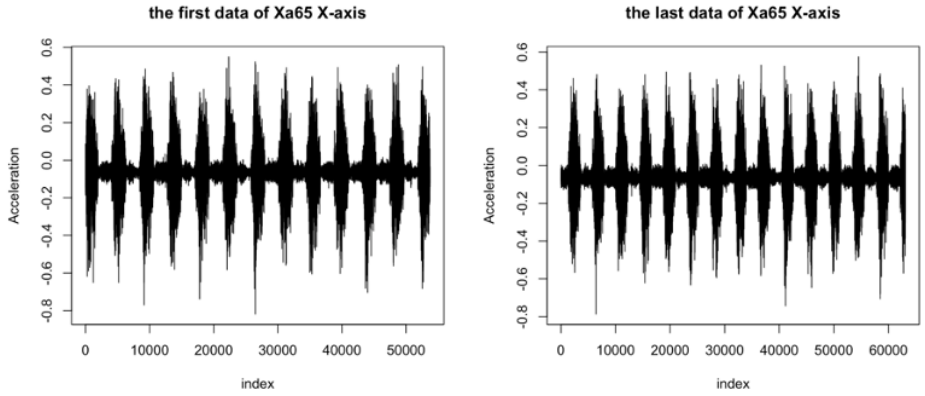

Fig 12: Vibration Signal Plots of the Initial and Final Data Segments

Boxplots of Xa65 Data Divided into Beginning, Middle, and End Segments (as shown in Figure 13)

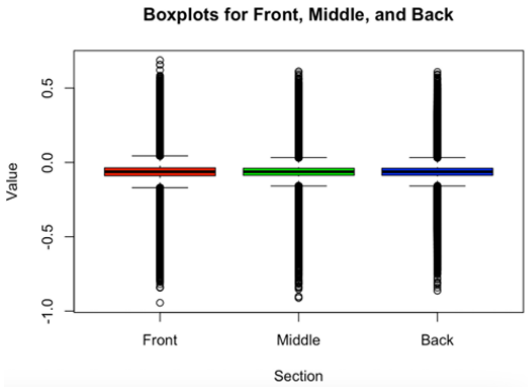

Fig 13: Boxplots of Xa65 Data Divided into Beginning, Middle, and End Segments

Based on the two figures above, no visually observable degradation can be detected; therefore, it is concluded that there is no degradation present. Moreover, the data background indeed corresponds to the five-minute vibration signals of the robotic arm, making it reasonable to determine that this dataset does not exhibit degradation.

5.2 Filter Out Static Signals

Each operation of the robotic arm is accompanied by a period of inactivity, which we refer to as the **non-vibration signal period** (highlighted in Figure 14). The non-vibration signal can interfere with the final determination of whether the robotic arm is in a normal or abnormal state. Therefore, it is necessary to remove the data corresponding to the non-vibration signal period.

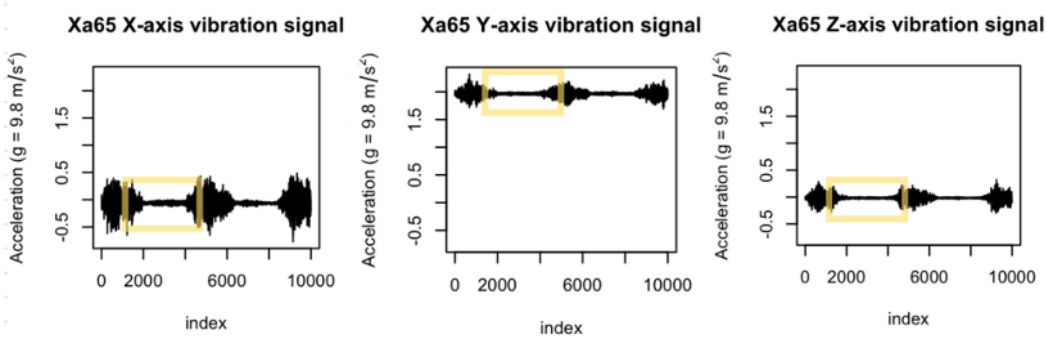

Fig 14: Three-Axis Vibration Signal Plot of Xa65

Next, we will attempt to remove the non-vibration signal using three different methods, including a windowed variance threshold method and two approaches for detecting the start and end points of the waveform. The advantages and disadvantages of each method will be discussed separately.

(1) Windowed variance threshold method

The absolute values of the X-, Y-, and Z-axes were summed and denoted as abs. A moving window of the first three and last three samples was applied to calculate the variance of abs within each window (as shown in Figure 15). A threshold was then set, and if the variance within a window falls below this threshold, the corresponding values are considered to represent periods when the robotic arm is stationary and are therefore removed.

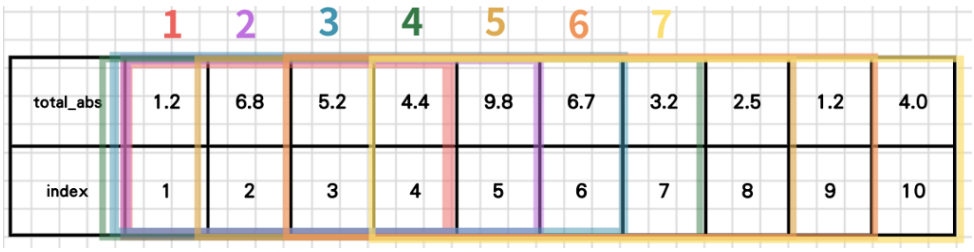

Fig 15: Illustration of Variance of abs within a Moving Window

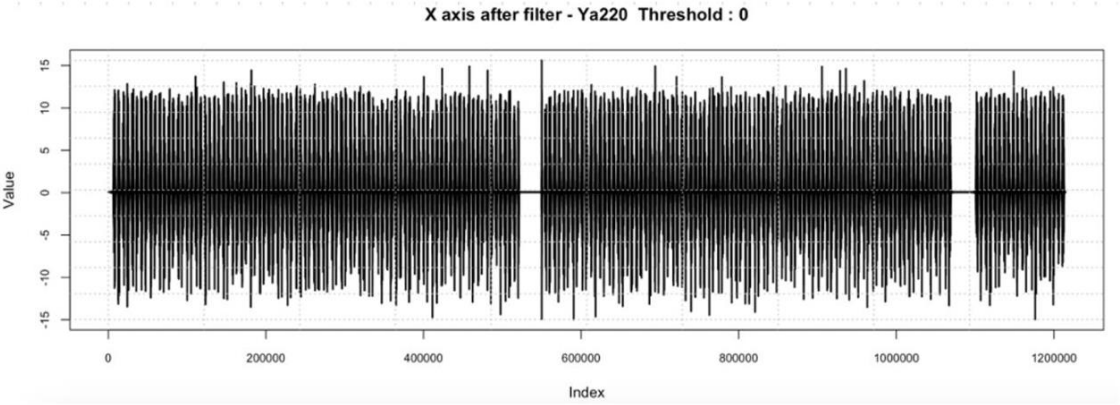

Fig 16: Vibration Signal Plot Before Removing Stable Segments Using Variance Threshold (Example: Ya220)

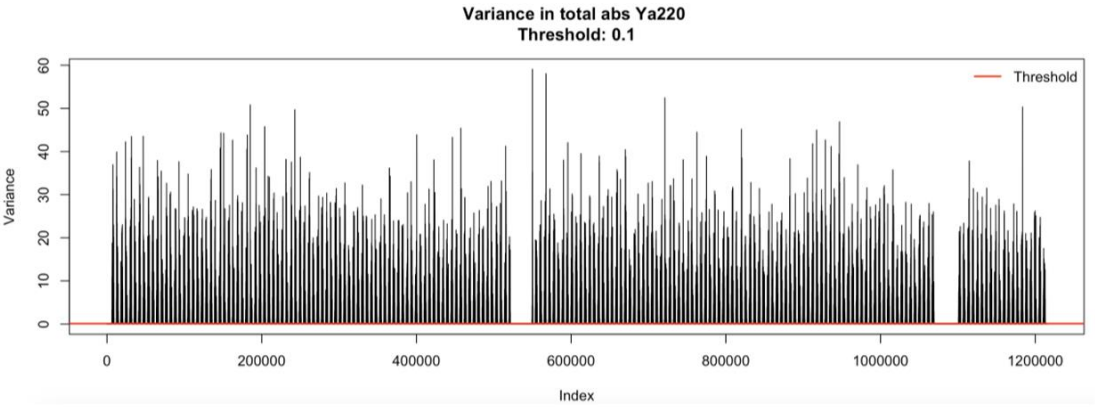

Fig 17: Example of Variance Threshold

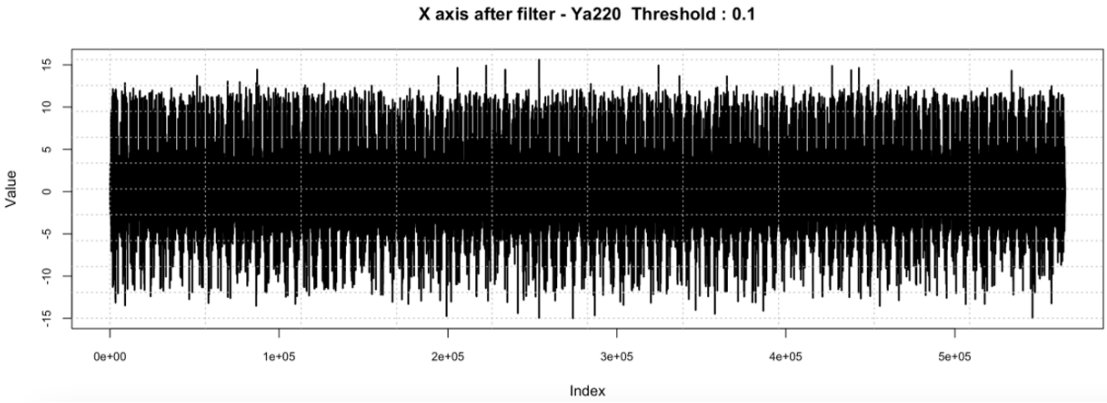

Fig 18: Vibration Signal Plot After Removing Stable Segments Using Variance Threshold (Example: Ya220)

Advantages: High computational efficiency and fully automatable.
Disadvantages: There is a risk of removing values within the waveform. For example, if the values within a wave happen to all be at the peak, resulting in minimal variation, the variance may still fall below the threshold and these values could be mistakenly removed.

(2) Detecting the start and end of the waveform – concept 1

For each axis (X, Y, and Z), the range of stable segments is first identified individually. Based on these ranges, the start and end points of the waveform are then determined.

Load / Position	65	80	95	130
Xa$X	-0.2 ~ 0.1	-0.2 ~ 0.1	-0.2 ~ 0.1	-0.2 ~ 0.1
Xa$Y	1.9 ~ 2.05	1.9 ~ 2.05	1.9 ~ 2.05	1.9 ~ 2.05
Xa$Z	-0.08 ~ 0.08	-0.08 ~ 0.08	-0.08 ~ 0.08	-0.08 ~ 0.08
Xb$X	-2 ~ -1.2	-2 ~ -1.2	-2 ~ -1.2	-2 ~ -1.2
Xb$Y	-1.5 ~ 1.5	-1.5 ~ 1.5	-1.5 ~ 1.5	-1.5 ~ 1.5
Xb$Z	-0.5 ~ 0.5	-0.5 ~ 0.5	-0.5 ~ 0.5	-0.5 ~ 0.5

Table 1: Vibrating range of Xa and Xb in static statement

Load / Position	220	260	300	380
Ya$X	-3 ~ 3	-3 ~ 3	-3 ~ 3	-3 ~ 3
Ya$Y	-2 ~ 2	-2 ~ 2	-2 ~ 2	-2 ~ 2
Ya$Z	1 ~ 3	1 ~ 3	1 ~ 3	1 ~ 3
Yb$X	-2 ~ 2	-2 ~ 2	-2 ~ -1.2	-2 ~ -1.2
Yb$Y	-1.5 ~ 1.5	-1.5 ~ 1.5	-1.5 ~ 1.5	-1.5 ~ 1.5
Yb$Z	-0.5 ~ 0.5	-0.5 ~ 0.5	-0.5 ~ 0.5	-0.5 ~ 0.5

Table 2: Range of static signals for each axis under different load levels at Ya and Yb

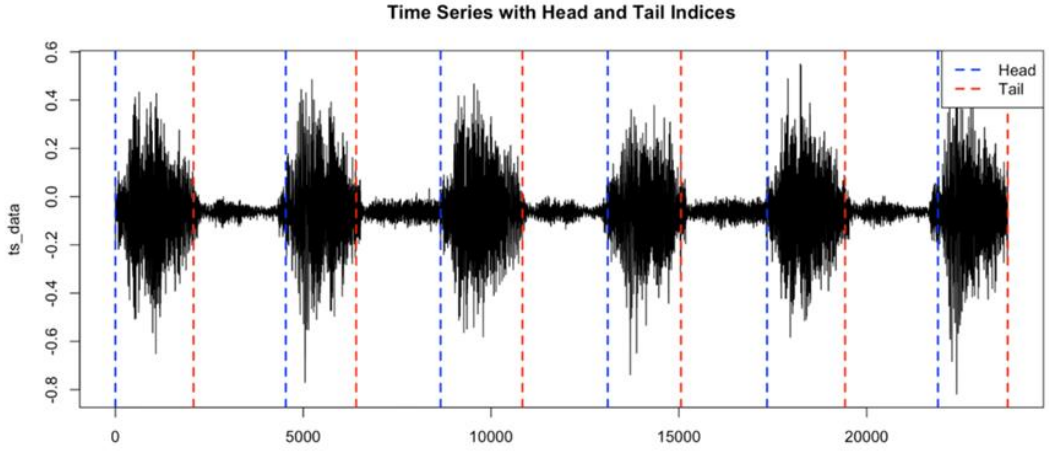

Fig 19: Illustration of waveform segmentation

gvb000001_X_ht

Segment_Head	Segment_Tail	Index	Data
1	1903	1	-0.17139538261847
1	1903	2	-0.23974250955062

gvb000001_Y_ht

Segment_Head	Segment_Tail	Index	Data
1	1912	1	1.92030137031769
1	1912	2	1.88935298080625

gvb000001_Z_ht

Segment_Head	Segment_Tail	Index	Data
1	1533	1	-0.0377485709666376
1	1533	2	-0.0811374044657039

Fig 20: Output data format by method (2)

Advantages: Ensures that values within the waveform are not removed, facilitates subsequent Fourier transform analysis, and enables more precise identification of vibration periods based on each axis.
Disadvantages: Due to inconsistent waveform lengths, automation is less efficient, and it is necessary to determine the range of non-vibration signals for each axis under different load conditions and sensor locations.

(3) Detecting the start and end of the waveform – concept 2
The range of the non-vibration segment is first identified based on the X-axis, and the start and end points of the waveform are determined accordingly. The Y- and Z-axes then adopt the same start and end indices obtained from the X-axis to identify their corresponding waveform boundaries.

gvb000001_ht

Segment_Head	Segment_Tail	Index	Xaxis	Yaxis	Zaxis
1	1903	1	-0.17139538261847	1.94027547943536	-0.0158814856771068
1	1903	2	-0.23974250955062	1.88446674750283	-0.0260903061480557
1	1903	3	-0.22841744103319	1.8851613387312	-0.0471376750625477
1	1903	4	-0.110896315477446	1.93232799168252	-0.0378148184417492
1	1903	5	0.0104970030885436	1.98218778405081	-0.0207927941155563
1	1903	6	0.0952942072288055	2.03958797375027	-0.0226872748073539
1	1903	7	0.0265493187502443	2.03621273426513	-0.0247143771592535
1	1903	8	-0.0553261931121386	2.02126391556015	-0.0202879986011843
1	1903	9	-0.0555001017690572	2.02006911624497	-0.0202466503185839
1	1903	10	-0.176590559583117	1.93714681160767	-0.00756931381864525

Fig 21: Fig 20: Output data format by method (3)

Advantages: Ensures that values within the waveform are not removed, facilitates subsequent Fourier transform analysis, maintains consistent waveform lengths, and requires only the non-vibration range from a single axis.
Disadvantages: Automation efficiency is relatively low, as it is still necessary to

determine the non-vibration signal range for each sensor location and load condition.

6 Introduction to the Analytical Methods

6.1 Entropy

According to the reference *Evaluation of Entropy Analysis* by Y. Jaen-Cuellar et al. (see page 28), which also focuses on fault diagnosis, the authors approach fault identification from the perspective of entropy. They argue that entropy-based methods are more effective in extracting information from vibration signals, as entropy can quantify the uncertainty and diversity of a signal.
In this study, six types of entropy are employed. **Spectral Entropy** is computed based on the energy distribution across frequencies obtained through Fourier transform. **Rényi Entropy** introduces an alpha parameter to adjust the weighting between high- and low-probability events. **Permutation Entropy** measures signal complexity by evaluating the permutations of subsequences. In contrast, **Approximate Entropy**, **Sample Entropy**, and **Fuzzy Entropy** incorporate a moving-window concept to quantify the similarity and regularity of the waveform.

6.2 Fast Fourier Transform

The Fourier transform is a method for converting time-domain data into frequency-domain representations and is widely used in signal analysis and detection. This technique decomposes a signal into a combination of sinusoidal and cosine waves at different frequencies and records the amplitudes of these components within the original waveform. By constructing a frequency spectrum, many detailed characteristics of the original signal can be observed.
Based on the nature of the data, Fourier transforms can be classified into continuous and discrete forms. As this study focuses on discrete data, the **Discrete Fourier Transform (DFT)** is applicable. Therefore, the **Fast Fourier Transform (FFT)**, which is derived from the DFT, is adopted as a preprocessing method to significantly reduce computational complexity and resource consumption.

6.3 Extract Segment-Wise Maximum

As the classification accuracy achieved by the original feature extraction methods still

had room for improvement, alternative feature extraction techniques were explored. Given that the degree of variation differs across segments of the frequency spectrum, a segmented feature extraction approach was proposed.
Since peak amplitude is a fundamental spectral feature that does not require additional parameter tuning or numerical transformation, and considering the importance of Fourier-based features, the highest peak amplitudes from the four sensor locations were identified as particularly informative. Therefore, segmented maximum peak amplitude extraction was selected as the initial approach.

7 Analysis and Validation Results

7.1 Entropy

After reviewing the literature and conducting experiments, we found that the machine analyzed in the referenced study is a **DC motor**, for which the returned signals are primarily voltage and current pulse signals, and rotational speed is a critical factor. In contrast, for robotic arms, **precise load control** is of greater importance, and system monitoring mainly relies on signals returned by sensors.
In addition, the experimental designs differ substantially. For example, the variables considered in the referenced study include the operating frequency of the DC motor and the gearbox speed, whereas this study focuses solely on the **load levels applied to the robotic arm**. As a result, entropy-based feature extraction did not achieve the expected performance in this context. Therefore, the following section introduces the hypothesis testing methods used in the analysis, including the **Kruskal–Wallis test** and the **Fisher Discriminant Score (FDS)**.
The **Kruskal–Wallis test** is a non-parametric statistical test that examines whether four groups of data originate from the same population by comparing the medians of ranked values. This test requires the samples to be independent but does not assume a normal distribution. The results of this analysis indicate that nearly all features reject the null hypothesis that data under different load conditions come from the same population.
The **Fisher Discriminant Score** is used to evaluate the effectiveness of features in distinguishing between different classes. By computing the within-class and between-class scatter matrices, FDS measures the discriminative power of each feature. Typically, an FDS value greater than 1 indicates that the feature is capable of separating different classes. In this study, the results vary across sensor locations: the performance at **Xa** and **Yb** is relatively poor, with only a small number of variables

achieving FDS values greater than 1, whereas **Xb** and **Ya** exhibit a larger number of variables with FDS values exceeding 1, indicating stronger discriminative capability for distinguishing different health conditions.

7.2 Fourier Transform

In this study, feature extraction was performed using the **Fast Fourier Transform (FFT)**. The segmented X-, Y-, and Z-axis vibration signals of each waveform were transformed from the time domain into the frequency domain, represented as combinations of sine and cosine waves. From the resulting spectra, the following fundamental features were extracted for each waveform:

1. **Variance, Quantiles, Mean, Median:** Descriptive statistics of the entire waveform.
2. **Dominant Frequency:** The frequency with the highest amplitude in the spectrum.
3. **Number of Peaks:** Peaks are defined where the amplitude difference between a frequency and its neighboring frequencies exceeds half of the range between the maximum and minimum amplitudes. The total number of such peaks is recorded.
4. **Peak Amplitude:** The maximum amplitude among all identified peaks.
5. **Spectral Energy:** Calculated as the sum of the squared amplitudes, representing the total energy of the waveform.
6. **Spectral Entropy:** Measures the complexity and irregularity of the spectrum; higher values indicate a more uniform distribution of frequency components.
7. **Frequency Bandwidth:** The range of frequencies containing the majority of the spectral energy.
8. **Frequency Centroid:** The "center of mass" of the signal's energy distribution in the frequency domain.
9. **Spectral Kurtosis (Sharpness):** Measures the peakedness of the spectrum, reflecting the concentration of frequency components.

var_x	median_x	q1_x	mean_x	q3_x	dominant_frequency_x	number_of_peaks_x
5778.648326	9.900153	5.750465	15.441124	15.214783	0.0	0
5981.544065	9.186477	5.865434	14.473980	13.670499	0.0	0
5392.452547	9.289726	5.845974	14.713276	13.572046	0.0	0
6434.370071	9.213310	6.050526	14.745507	14.399582	0.0	0
6227.298020	10.137979	6.323567	15.498301	15.282176	0.0	0

peak_amplitude_x	spectral_energy_x	spectral_entropy_x	bandwidth_x	spectral_centroid_x	spectral_kurtosis_x
1673.033105	2.912265e+06	7.405788	0.0	0.187943	0.965094
1765.808187	3.219341e+06	7.427332	0.0	0.175950	0.972398
1577.802187	2.585718e+06	7.300492	0.0	0.181211	0.965968
1884.528535	3.665142e+06	7.514677	0.0	0.181756	0.972415
1810.577678	3.401903e+06	7.535032	0.0	0.188921	0.967760

Fig 22: Dataframe format

After extracting the above features for the X, Y, and Z axes, each axis contains all of these features. To reduce model complexity and improve interpretability, **Fisher feature selection** was employed to identify the most significant features at each sensor location.

Features	Xa	Xb	Ya	Yb
Variance		√	√	
Quantiles		√	√	
Mean		√	√	
Medians		√	√	
Dominant Frequency			√	
Number of Peaks			√	
Peak Amplitude		√	√	
Spectral Energy		√	√	
Spectral Entropy		√	√	
Frequency Bandwidth		√	√	√
Frequency Centroid		√	√	
Spectral Kurtosis		√	√	√

Fig 23: Significance of Fisher Feature Selection

It can be observed that the features from **Xa** and **Yb** exhibit relatively weak discriminative power for distinguishing health conditions, whereas the features from **Xb** and **Ya** perform considerably better. To prevent excessive imbalance during subsequent model training, all features were retained and included in the model.

Model selection: Random Forest

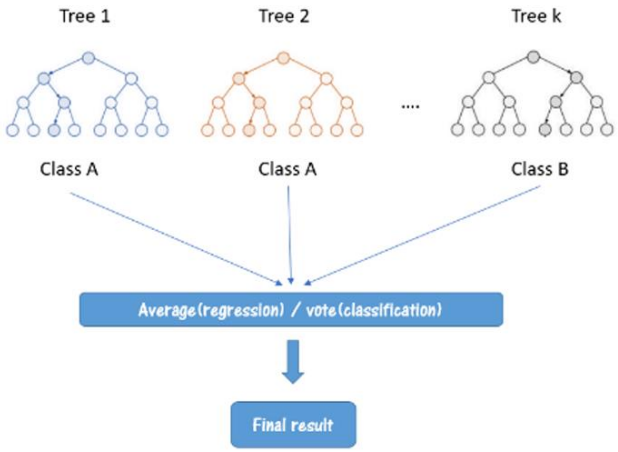

Fig 24: Random Forest structure

Considering the large number of feature variables and the uncertainty regarding potential interactions among the X, Y, and Z axes, the **Random Forest** method was employed. Random Forest can efficiently explore interactions between axes, and since each decision tree uses a randomly selected subset of features, this approach increases model diversity, reduces the risk of overfitting, and mitigates the effects of multicollinearity.

Position	Accuracy (%)
Horizontal Motion – Motor Side (Xa)	88.13
Horizontal Motion – Idler Side (Xb)	92.81
Vertical Motion – Motor Side (Ya)	99.18
Vertical Motion – Idler Side (Yb)	98.36

Table 3: Model results

The classification results of the Random Forest show that, except for the **Xa** location, all other sensor positions achieved an accuracy of over 90%. In terms of load direction, the classification performance for **vertical motion** is noticeably better than that for **horizontal motion**.

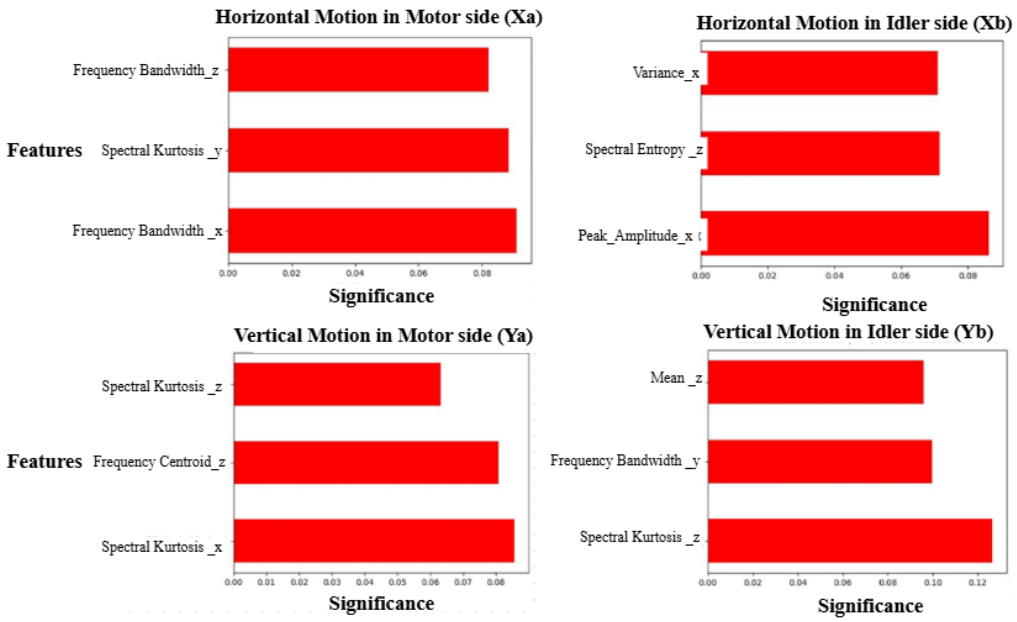

Fig 25: Top 3 significant features in each position

Based on the feature importance shown in the figure above, the most important variables at each sensor location are as follows: **Xa** – X-axis frequency bandwidth, **Xb** – X-axis peak amplitude, **Ya** – X-axis spectral kurtosis, and **Yb** – Z-axis spectral kurtosis. The top three features at the **Yb** location account for a higher proportion of importance compared to the other three locations. Although no clear pattern is apparent, these results can still provide guidance for subsequent feature extraction and allow for feature selection adjustments based on sensor location.

7.3 Extract Segment-Wise Maximum

The figure below illustrates the **segmented maximum peak extraction** (Figure 26). Using 9 segments as an example, the red dots indicate the maximum value in each segment.

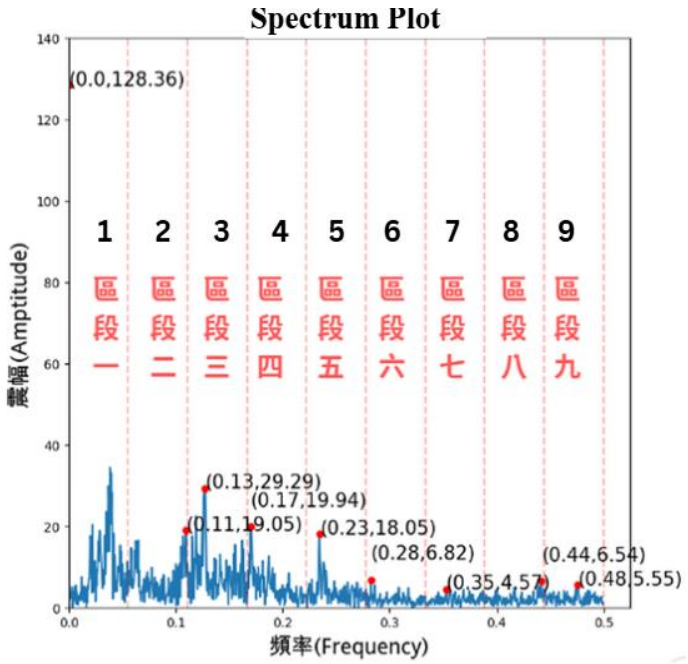

Fig 26: Illustration of Segmented Maximum Peak Extraction
Using 9 segments as an example, the X, Y, and Z axes each have their respective segment-wise maximum values, as shown in Figure 27

此波頭 Index	此波尾 Index	X軸-區段一	Y軸-區段一	Z軸-區段一	頻率種類	X軸-區段二	Y軸-區段二	Z軸-區段二	X軸-區段三	...	Z軸-區段六	X軸-區段七	Y軸-區段七	Z軸-區段七	X軸-區段八	Y軸-區段八	Z軸-區段八	X軸-區段九	Y軸-區段九	Z軸-區段九
1	2078	128.357992	4083.127587	36.099974	Xa130	19.045750	14.166618	24.695610	29.293866	...	2.626262	4.569690	2.585464	2.276482	6.541043	4.089064	1.928640	5.552386	3.614318	2.201766
4544	6481	121.941518	3808.073081	33.078000	Xa130	22.251669	14.890844	11.624025	28.918423	...	2.159985	3.679214	2.256582	2.044124	4.350336	2.907558	1.879337	5.399984	2.286200	1.655007
8804	10760	124.393957	3846.764284	34.911610	Xa130	21.301692	15.989202	14.895162	28.040685	...	2.373301	4.212958	1.945391	2.036227	5.331618	3.091522	2.107222	5.276813	3.399689	1.956968
13094	15170	128.994947	4082.883088	37.294364	Xa130	15.290998	14.416036	17.967302	35.133458	...	2.343268	3.607413	2.282151	1.994728	4.559290	2.539331	1.571029	5.844351	2.654912	1.457851
17359	19418	128.390287	4051.190050	36.668222	Xa130	26.851615	12.316387	25.209910	32.410869	...	2.838323	3.770598	3.518398	2.070583	4.481467	2.291585	2.510453	5.588660	2.445783	2.222841
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1162004	1163830	115.424328	3588.981654	31.916747	Xa95	20.141984	14.144668	17.382877	25.162152	...	1.924958	3.752482	3.351303	1.903265	4.090498	2.420726	1.820957	5.394812	2.159510	1.716154
1166122	1168303	139.745859	4287.414459	36.515553	Xa95	17.841407	11.189808	23.785653	19.740897	...	3.087204	4.209344	2.340858	1.666500	5.423210	2.946645	2.374184	4.901717	2.058030	1.610647
1170755	1172536	110.915780	3503.886184	30.745418	Xa95	18.596026	13.839569	15.658039	21.673802	...	1.902451	3.396531	2.366184	1.575459	5.533193	2.969327	1.682535	4.251967	2.383180	1.629175
1174811	1176991	139.317341	4286.819476	38.242303	Xa95	14.163586	9.395683	17.898269	24.944048	...	2.675327	3.988610	2.243270	2.118932	5.103159	2.739998	1.975428	5.673748	2.087456	1.894744
1179466	1181228	116.107926	3465.299783	31.688821	Xa95	21.851793	17.169592	20.096285	22.535663	...	1.735784	3.177856	2.283248	1.580838	4.137297	2.420999	1.555832	4.088187	2.446283	1.422006

Fig 27 : DataFrame Format

However, we found that the optimal number of segments may differ for each sensor location. Therefore, for each location, a fixed training set and test set were first selected, and the performance was evaluated across different numbers of segments. Figure 29 shows the accuracy line plots for all four sensor locations.

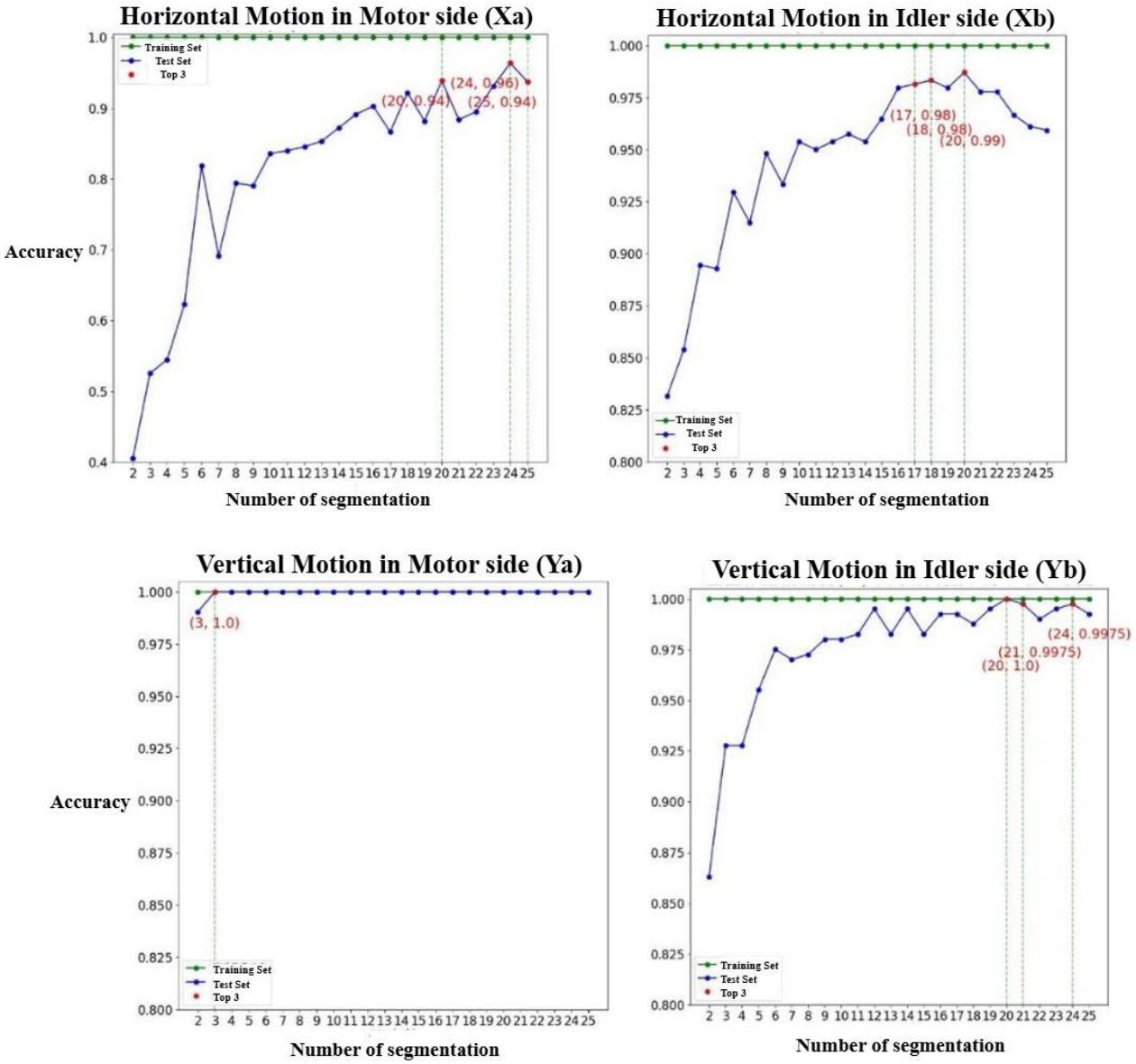

Fig 28: Accuracy at Each Sensor Location under Different Numbers of Segments

The red dots indicate the number of segments that we consider to yield better performance for each sensor location. Cross-validation was then used to select the segment number that provides more stable accuracy, as shown in Figure 29.

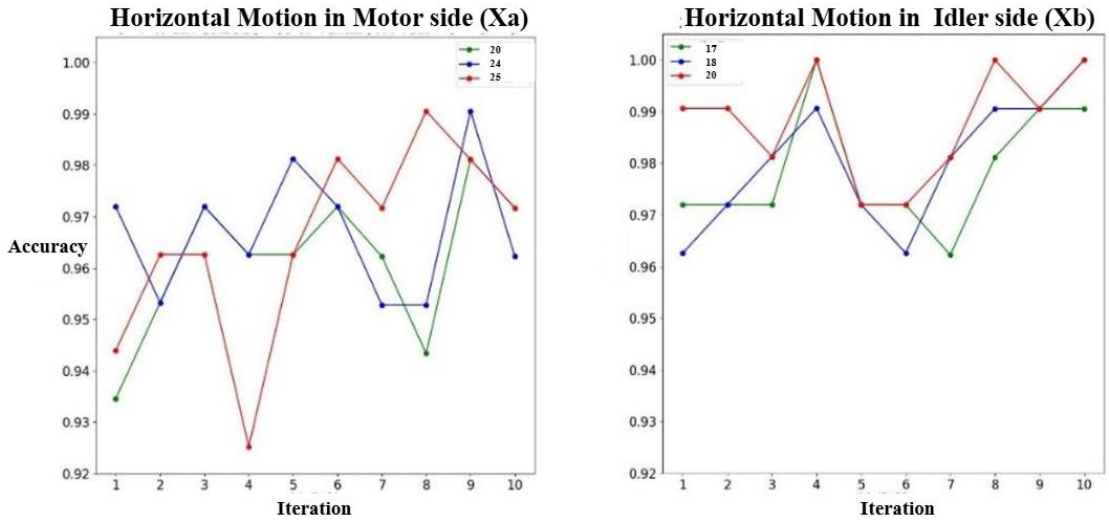

Fig 29: Cross-validation results under different numbers of segments

For horizontal motion, a segment number of **24** was chosen for the motor side and **20** for the idler side, as these values provided more stable and higher performance. Additionally, when detecting load anomalies in horizontal motion, installing the sensor on the idler side is preferable to the motor side, since it achieves higher accuracy.

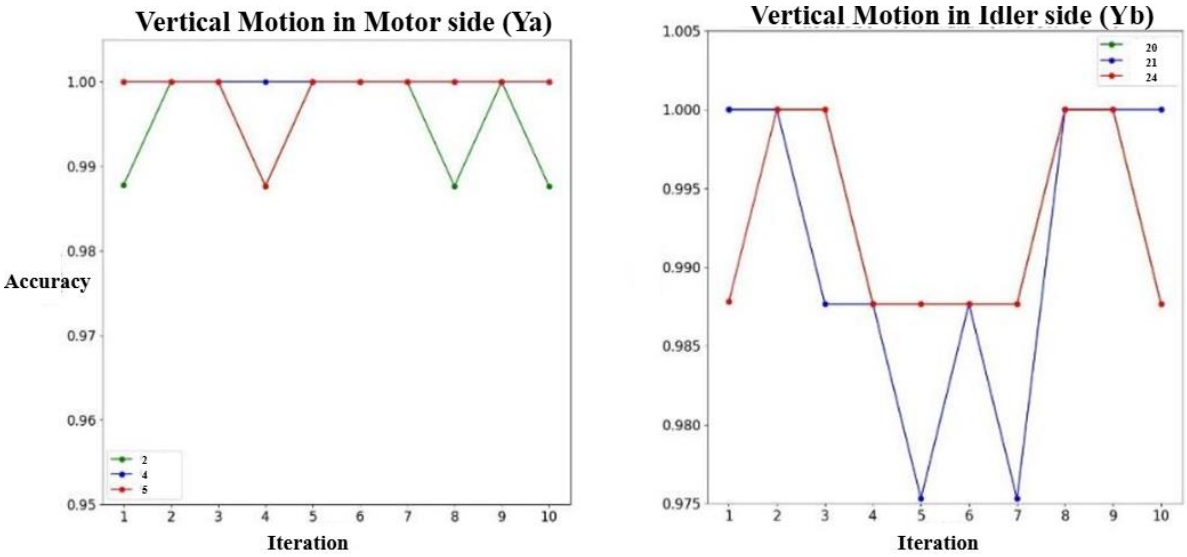

Fig 30: Cross-validation results under different numbers of segments

For vertical motion, a segment number of **4** was chosen for the motor side. On the idler side, the results for segment numbers **24** and **20** were similar, so **20** was selected to reduce computational load. Furthermore, when detecting load anomalies in vertical motion, installing the sensor on the motor side is preferable to the idler side, as it provides higher and more stable accuracy.

8 Design of the Decision Support Tool

The design concept assumes that users of this auxiliary tool are all in possession of robotic arm vibration data. Upon entering the interface, users must first select the motion location they wish to observe and control the number of files to upload. Due to memory limitations in **R Shiny**, the maximum number of uploadable files is set to 25. After clicking **Process Files**, the uploaded data is sent to the server for preprocessing, **Fast Fourier Transform (FFT)**, feature extraction, and model evaluation.

Fig 31:

Based on the initially selected motion location, the dashboard is designed to display a **pie chart** in the lower-left corner, showing the proportion of each load at that location. Healthy loads are indicated in green, while all others are shown in red (tentative). In the upper-right corner, the dashboard will display the overall proportion of healthy data in the uploaded files, as well as the individual counts of healthy and unhealthy files.

To allow users to better understand the health status of each uploaded file, the lower-right section is designed as a **health overview table**. When the user hovers the cursor over each cell, the file name and corresponding load will be displayed. The blank space at the top is reserved for information such as the next model update time, customer service phone number, and relevant personnel emails.

From the user's perspective, we recognize that not all users have a background in model building or statistics; their main interest is simply the health status of the uploaded data. Therefore, the dashboard will **not display raw data or vibration signals of waveform heads and tails**. Only clear and concise information about health status and whether the arm is functioning properly will be presented.

Additional Notes:

1. If the user selects a motion location that does not match the actual location of the uploaded files, a pop-up error message will appear.
2. The dashboard can be used repeatedly without needing to refresh the page.

To ensure user confidence in the model's accuracy, the following basic after-sales services are provided:

1. **Accuracy Guarantee:** If the accuracy falls below 95%, an alert email will be sent to the company, which will then dispatch technical personnel for updates or maintenance.
2. **Customized Models:** The model will be continuously trained with customer data to improve accuracy, ensuring users can trust the auxiliary decision-making tool.

For users who purchase the **VIP plan**, the following additional services are provided:

two months of periodic model checks and updates, a detailed health report, and priority support for troubleshooting issues.
For more details on dashboard implementation and operation, please refer to the following website:
[GitHub - Iris910531/BigData_course]

9 Conclusion

Among the three methods discussed and validated earlier (entropy, Fast Fourier Transform, and segmented maximum peak extraction), **entropy** was ultimately not adopted due to unsatisfactory results in the hypothesis testing. Currently, two feature extraction methods remain for selection. Our goal is to achieve stable results with high accuracy, so **cross-validation comparisons** were conducted for each sensor location using the two methods. The results are shown in the figure below.

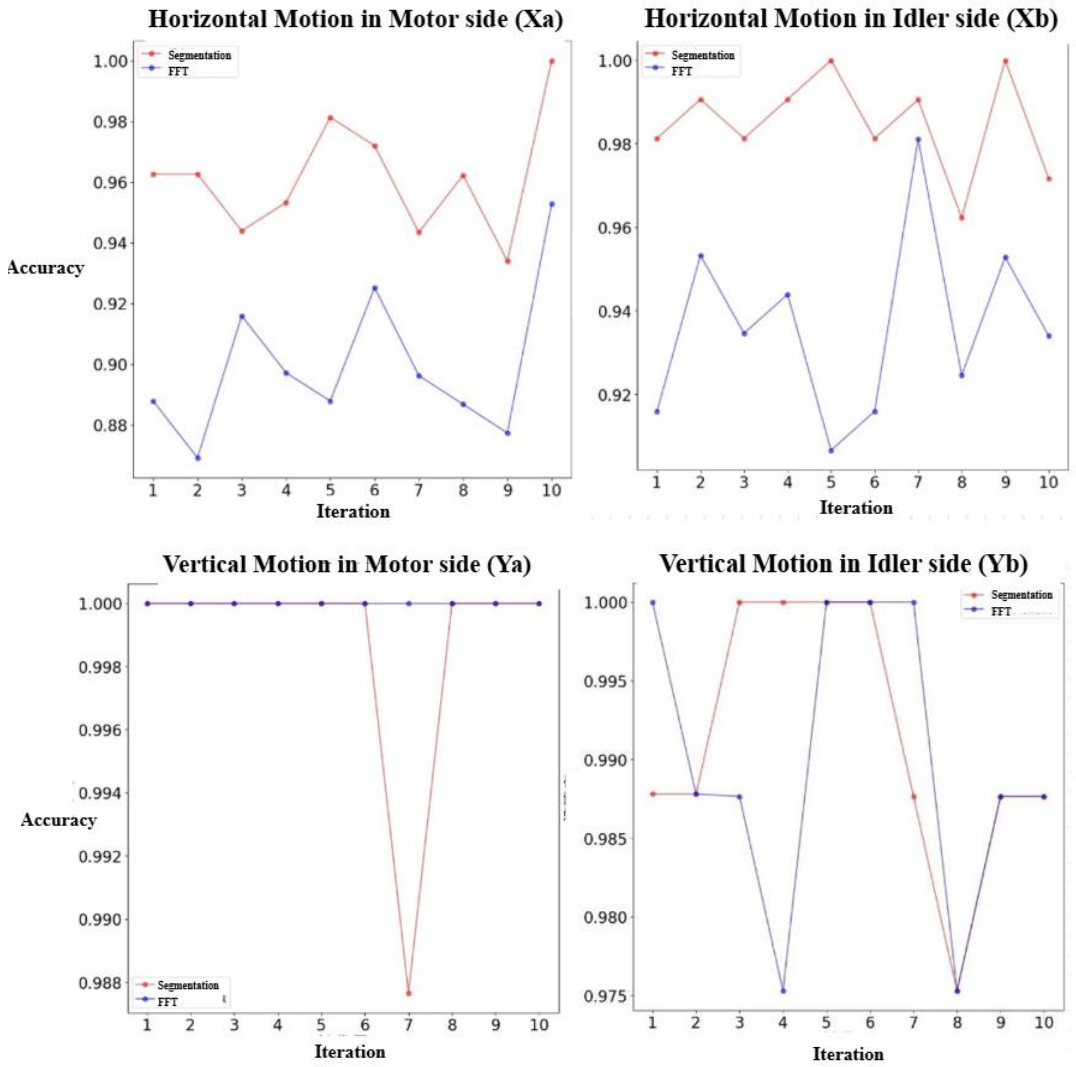

Fig 32: Accuracy comparison between two methods (Segmentation vs. FFT)

From the figure, it can be observed that the **segmented maximum peak extraction** method generally performs better. To ensure consistency across all prediction methods, we ultimately selected this method for feature extraction.
Based on the accuracy results under different segment numbers, **Xb** consistently shows higher and more stable accuracy than **Xa**, indicating that for horizontal motion, installing the sensor on the **idler side** is more suitable for anomaly detection.
Similarly, since **Ya** demonstrates better accuracy than **Yb**, the sensor for vertical motion is more suitably placed on the **motor side**.
We believe that models trained using the segmented maximum amplitude peaks achieve higher performance than models trained with all features. This is primarily because segmentation allows each waveform to be divided more finely, enabling more precise model training. In the future, we aim to use **Fisher feature selection** to identify the most important variables and apply segmented peak extraction on these variables during model training, with the goal of obtaining a more stable and higher-performing model.
For this project, we suggest that providing additional background information, such as experimental videos or flowcharts, could help identify more appropriate reference literature. Furthermore, for the experiments, increasing the variety of load conditions could make the model more robust and broadly applicable.

10 Team Members and Responsibilities

Project Manager	● 侯言蓁: ■ Gantt Graph ■ Presentation preparation ■ Meeting records ■ Final report preparation ● 戴衣伶: ■ Presentation preparation ■ Flowchart design
	● 林瑋珈: ■ Fourier Trasnformation ■ Feature Extraction ● Yi-Hsuan Kuo: ■ Data preprocessing ● 薛珮妤: ■ Statistical inference for

Data Scientist	entropy ● Bo-Ye, Li: ■ Entropy-based feature extraction ■ Literature Review about entropy ■ Model training ● 陳皓鈞: ■ Literature Review about Fourier transformation ■ Model training
System Designer	● 戴衣伶: ■ Dashboard design and implementation ● Yi-Hsuan Kuo: ■ Dashboard design and implementation ■ Push code to Github ● 陳皓鈞: ■ Dashboard implementation ● 薛珮妤: ■ Dashboard implementation

11 Team Members' Contributions

- **Yi-Hsuan Kuo:**

  In this project, I was responsible for automating the preprocessing of vibration signals to effectively remove noise and improve data accuracy and reliability. In addition, I organized and led multiple project meetings to ensure smooth collaboration and steady progress within the team. During these meetings, I also took charge of documenting key decisions and action items.

  Regarding dashboard design and implementation, I led the entire design process and was responsible for the actual programming implementation. The dashboard integrates preprocessed data, model applications, and final results into a unified and intuitive visualization platform. Furthermore, I was responsible for

integrating all source code, including data preprocessing, model deployment, and result presentation, to ensure overall consistency and efficiency of the project.

Finally, I uploaded the complete dashboard implementation, along with environment configurations and package compatibility details, to GitHub. This enables easy access for other team members and future users, ensuring the project's sustainability and scalability.

- **Bo-Ye, Li:**
  In this project, I was involved in **data visualization**, **data preprocessing**, **feature extraction**, and **model training**.

  First, during the data visualization stage, I analyzed vibration differences under varying load conditions across different locations. Next, to remove idle-state segments, I applied a moving-window variance calculation and filtered out low-variance portions of the signal.

  Subsequently, based on relevant literature, I performed feature extraction using six types of entropy measures. After applying the Fast Fourier Transform (FFT), the resulting frequency spectra were segmented, and maximum values were extracted. The optimal number of segments for each location was then selected.

  Finally, a Random Forest model was trained, and K-fold cross-validation was employed to ensure model stability. The final model achieved an accuracy exceeding 95%.

- 戴衣伶:
  In this project, I was primarily responsible for **presentation preparation and visual design, project execution planning and flowchart design**, as well as **the design and partial implementation of the dashboard using R Shiny**.

  For the presentations, I carefully designed and produced professional slides to ensure clear communication of key information while enhancing visual appeal. In terms of project planning and flowchart design, I systematically outlined and documented each project phase and workflow, providing a solid foundation for successful execution and ensuring an organized and efficient working process.

  Additionally, I designed and contributed to the partial implementation of the dashboard using R Shiny, enabling real-time visualization of key data. This ensured data accuracy and timeliness, providing strong analytical support for

informed decision-making. Overall, my contributions significantly improved both the efficiency and effectiveness of the project and laid a solid foundation for its success.

- **薛珮妤:**
  In this project, I conducted data visualization using boxplots to examine data distributions, which helped identify preprocessing criteria and improve the accuracy of data selection. Prior to model building, I performed a thorough review of relevant literature and studied Fourier Transform techniques to scientifically select the most informative features for classification and prediction using rigorous statistical tests.

  Finally, I developed an interactive dashboard using R Shiny, enabling dynamic data visualization and analysis. This enhanced both the professionalism and practicality of the project, allowing users to conveniently explore and analyze data, thereby increasing the project's applied value and improving user experience.

- **陳皓鈞:**
  We conducted a comprehensive literature review on vibration data analysis, with a particular focus on the application and limitations of the Fourier transform as a foundational component of our project. After my teammates successfully extracted the fundamental characteristics of the vibration waveforms, we built a random forest model incorporating all extracted features. This model not only deepened our understanding of the data but also served as a key benchmark for comparing subsequent models. During the implementation of the dashboard, I worked closely with my teammates and provided ongoing moral support to ensure that the visualization interface accurately and completely represented the results of our research.
- **林瑋珈:**
  In this project, I first developed a thorough understanding of the principles of the Fast Fourier Transform (FFT) and its wide range of applications, ensuring a solid theoretical foundation and practical flexibility. Next, I transformed the preprocessed data into the frequency domain, laying the groundwork for subsequent analysis and model training.

For the robotic arm application, I conducted an extensive literature review to extract a large number of valuable features, which were crucial for training efficient models. I completed the feature extraction process for all datasets, making significant contributions to the project's success and enabling high accuracy during model training.

Additionally, my reporting and writing skills were fully applied in both presentations and written reports, further enhancing the overall quality and impact of the project.

- 侯言蓁:

  In this project, I meticulously documented weekly meeting notes and compiled questions for consultation with the professor, ensuring timely guidance and feedback. During the first presentation, I explained the overall project direction and objectives, and introduced preliminary data observations, laying the foundation for subsequent analyses and reports.

  Additionally, I created a Gantt chart to plan project tasks and timelines, ensuring the project progressed according to schedule. Across three presentations, I was responsible for slide design and formatting, ensuring clarity and visual appeal to enhance communication effectiveness.

  I also consolidated the team's dispersed results and took charge of the final written report, including content drafting, layout design, and proofreading, ensuring completeness and professionalism. Through these responsibilities, I demonstrated strong project management skills and team collaboration, contributing significantly to the smooth execution of the project and substantially improving the quality and impact of our reports.

12 Meeting Minutes

12.1 Weekly meeting minutes

03/30

1. Literature Review

- Discussion on the **feasibility of the PHM 2012 dataset**

- Covered topics include:
    - **Feature engineering**, such as skewness and kurtosis
    - **Remaining Useful Life (RUL) estimation**
    - **Principal Component Analysis (PCA)** (with concerns about computational cost)
    - **Fast Fourier Transform (FFT)**

2. Research Direction

- Initial focus on:
    - **Data exploration / data mining**
    - **Data reduction / dimensionality reduction**

3. Task Allocation

- Dimensionality reduction: **Yi-Hsuan Kuo**
- Exploratory Data Analysis (EDA): **Yi-Hsuan Kuo & Bo-Ye**
- Gantt chart preparation: 侯言蓁

4/5

1. Revision of Initial Assumptions

- The original approach was incorrect: **data collection and transmission require time**, therefore the sampling intervals in the .txt files may **not be uniformly distributed**.
- Periods with relatively stable acceleration values actually correspond to **machine idle states**, rather than meaningful pattern-related features.
- As a result, **data corresponding to machine idle states should be removed**.

2. Methods Considered for Identifying Idle States

1. **Local variance-based filtering**
    - Compute the variance using each data point and its **three preceding and three succeeding points**.
    - Points with very small variance (e.g., < 0.01?) are regarded as **machine idle** and removed.
    - Issue: examining **x, y, and z axes independently** may lead to inconsistent removal (e.g., x removed while y or z remains).
2. **Variance of distance (Euclidean norm)**
    - Use the variance of the **Euclidean distance** derived from x, y, and z.
3. **Moving average**
    - Concern: **sign-crossing issues** caused by positive and negative values.
4. **Merging x, y, z signals**
    - Sum the **absolute values** of x, y, and z, then apply a threshold.

3. Threshold Selection Discussion

- Should the **same threshold** be applied to x, y, and z?

1. Determine acceleration levels of the robotic arm that indicate actual motion.
    - Mean ± 3 standard deviations was considered, but this approach is **not practical** due to a large number of zeros and the fact that ±3σ still covers some non-idle data.
2. Merge x, y, z by summing absolute values and select a threshold.
3. Use variance of the Euclidean distance.
    - Concern: at vibration peaks, the difference between consecutive points may be small, potentially causing **valid data to be mistakenly removed**.

4. Final Decision

- Threshold-based methods were **not adopted**.
- Instead, **waveform head–tail extraction** was used to identify and retain meaningful segments.
- Time-series modeling methods may **not be applicable**; the analysis will primarily focus on **Fourier-based approaches (FFT)**.

5. Task Allocation

- Time-series plots: **Yi-Hsuan Kuo & Bo-Ye**
- Fourier analysis: 陳皓鈞 & 林瑋珈
- Visualization: 薛珮妤 & Bo-Ye
- PPT organization and layout: 侯言蓁 & 戴衣伶

4/22

1. Literature Review (Found by the DS Team)

- The papers reviewed by **Yi-Hsuan Kuo** and **薛珮妤** were considered **less applicable** to the current problem.
- The **entropy-based approach** reviewed by **Bo-Ye** was considered **promising and feasible**.

2. Methodological Ideas

- For **stationary signals**, apply the **Fast Fourier Transform (FFT)** followed by **denoising**.
- Identify the **start and end** of meaningful signal segments.
- Adjust segment length to a **power of two (2^n)**:
    - Either truncate the signal, or
    - Apply **zero-padding** at the beginning and/or end.

3. Task Allocation

- Data preprocessing: **Yi-Hsuan Kuo**

04/27

1. Explanation of Entropy-Based Literature and Methods

- The entropy-related literature and corresponding analysis methods were explained to the team.
- Due to **time constraints**, **entropy-based methods and Fourier-based methods may not both be fully implemented**.
- The team decided to **prioritize Fourier-based analysis**, with entropy analysis as a secondary option.

2. Fourier-Based Feature Considerations

- Features to be extracted from the Fourier domain include:
  - **Frequency**
  - **Amplitude / energy**
  - **Phase** (e.g., peak locations)

3. Task Allocation

- Data preprocessing (window segmentation): **Yi-Hsuan Kuo**
- Fast Fourier Transform (FFT): 陳皓鈞, 林瑋珈
- Entropy analysis: Bo-Ye, 薛珮妤
- Gantt chart revision and workflow diagram: 侯言蓁, 戴衣伶

5/5

1. Discuss Results from Last Week (Presented by the DS Team)

- The DS team explained the results obtained last week, including:
  - **Data preprocessing**
  - **Fourier-based analysis**
  - **Entropy-based analysis**

2. Presentation Preparation

- The **second presentation slides** were generally revised and refined.
- **Speaking roles and script assignments** were discussed and allocated.

3. Methodological Discussion

- A brief discussion was held on the **logistic regression approach**.

4. Task Allocation

- Correlation analysis / statistical tests: Pei

5/6

Prepare the second presentation

5/10

1. Feature Selection after Fourier Transform

- Discussion on which **features should be retained after FFT**, such as:
  - **Frequency bands / bandwidths**
- Consideration of how these frequency-domain features should be represented for modeling.

2. Modeling Approaches

- Three models will be implemented and compared:
  1. **Random Forest**
  2. **Support Vector Machine (SVM)**
  3. **Logistic Regression**
- Model training will involve **feature input and hyperparameter tuning**.

3. Dashboard Design Discussion

- Discussion on how to **remove steady-state (idle) segments** from customer data in the dashboard.
- Options considered:
  - Using **fixed thresholds**, or
  - Developing an **automated method**, which is preferred if feasible.

4. Task Allocation

- Testing and data handling: **Yi-Hsuan Kuo**
- Feature engineering: 林瑋珈
- Feature significance testing: 薛珮妤
- Feature visualization: Bo-Ye

5/18

1. Classification Performance

- **XA**: Classification performance was **not very strong**
- **XB**: Classes were **generally well separated**, with no major issues observed.

2. Modeling Considerations

- Discussion on how to handle **x, y, z axes**:
  - Should they be **modeled separately** (and discard less informative ones), or
  - **Combined** into a single representation?
- Models under consideration:
  - **Random Forest**
  - **Support Vector Machine (SVM)**
  - **Logistic Regression**

- Feature construction before model training:
    - Should x, y, z be merged using **squared distance (e.g., Euclidean norm)**, or
    - Using the **sum of absolute values** to form a single composite variable?

3. Dashboard Discussion

- The **PM** will act as a **challenging customer**, raise critical questions, and **document them in written form**.
- Instructor's suggestion:
    - Use **signal segmentation (slicing/windows)** to analyze and present the data more effectively.

5/25

1. Dashboard Test Data Input

- Default test input data works properly; however, the **input length must be sufficiently long**.
- Each .txt file typically contains **around four waveform segments**.

2. Whether to Display Vibration Signals

- **Arguments against displaying vibration signals**:
    - Customers may **not be interested** in viewing raw vibration signals.
    - As long as the **input data length is adequate for prediction**, visualizing signal head–tail segments is unnecessary.
    - Displaying such plots may only **increase user confusion or concern**.
    - Even if customers notice imperfect head–tail extraction, they may have **no practical way to fix it**.
    - As long as a prediction result is provided, the process should avoid causing customers to **doubt the model's reliability**.
- **Arguments for displaying vibration signals**:
    - Allows verification of whether **head–tail extraction is performed correctly**.
    - If some inputs are **too short for proper segmentation**, the system could advise customers to provide alternative data.
    - However, cases where only a single head–tail segment is extracted may **not necessarily indicate an anomaly**.

3. Model Selection and Parameter Settings

- Models discussed: **Random Forest** and **Logistic Regression**.
- Random Forest parameters:
    - Number of trees (default: **100**).
    - Whether hyperparameter tuning is necessary.
- Conclusion:

    - Customers are **not expected to understand model details**, so exposing such options is unnecessary.

4. Final Decision

- The system will **only adopt Random Forest** as the final model.
- Using multiple models could lead to **inconsistent results**, which may cause confusion or awkward situations.
- Therefore, the **model selection option will be removed** from the dashboard.

6/1

1. Segmentation Strategy and Performance

- The **entire waveform** was segmented into multiple slices.
- Approximately **2 to 25 segments** were examined, and the **classification accuracy was consistently high** across different segmentation settings.
- This indicates that **window-based segmentation is highly effective**.

2. Physical Configuration Insight

- It is hypothesized that placing the **motor vertically** and the **idler horizontally** leads to **better performance**.

3. FFT vs. Segmentation Comparison

- Results from **Fourier-based features** were compared with **segmentation-based features**.
- Overall, **segmentation-based methods performed better** than FFT-based approaches.

4. Head–Tail Extraction Revision

- The previously used head–tail (start–end) detection methods were **inconsistent across axes**, making it difficult to handle **interaction effects**.
- As a result, head–tail extraction was **standardized using the x-axis only** to segment the waveform.

5. Feature Selection Across Axes

- From a large set of variables, features that showed **clear patterns across XA, XB, YA, and YB** were selected.

Future Work

- Based on **feature importance**, additional important variables can be selected for segmentation.
- This may lead to **more stable and accurate models**.

6. Dashboard Design Decisions

- A **threshold-based rule** is applied:
    - Data with head–tail values **below 500** are treated as abnormal and are **removed**.
- A **segmentation-based approach** is used uniformly to handle abnormal data.

- The size of generated features is monitored, although it is **difficult to manage directly**.

7. File Upload and Error Handling

- The dashboard is designed to allow **a limited number of files per upload** (e.g., **25 files**) for easier configuration and processing.
- Informative messages will be printed:
  - Preprocessing includes **validation checks**.
  - Invalid data will **not be passed to subsequent models**.
  - Instead, the system will directly report **data abnormalities** at the final stage.
  - Abnormal files are identified by **file name**, and processing is organized by **batches of 25 files**.

8. Result Visualization

- Results will follow the instructor's example:
  - Use **colored square blocks** to represent outcomes.
  - Additionally present results using **pie charts**, **health indices**, and **abnormality rates**.

6/3

Revise the third presentation

6/10

Dashboard Additional Features

1. Information Display

- **Next model update time**
- **Customer service hotline**, **contact email**, and **company address**
- **Customer satisfaction feedback**
- **VIP subscription plan**:
  - Bi-monthly model updates
  - Health check reports
  - Priority handling of customer inquiries

2. Non-VIP / Basic After-Sales Service

- If model accuracy falls below a predefined threshold (e.g., **95%**), an **automatic Gmail notification** will be sent to the company.
- The company will **dispatch personnel to update and maintain the system** as needed.
- Customer-provided data will be **continually used to retrain the model** (customized training) to improve prediction accuracy and **increase customer trust** in dashboard results.

12.2 Q&A Log for Each Class Presentation

(I) April 9, 2024 – First Report Q&A

Instructor Questions & Team Responses

1. **Q:** Why is page 11 left blank?
   A: To keep the **y-axis fixed** for better comparison.
2. **Q:** On page 34, why do the X, Y, Z axes appear to show zero vibration?
   A: Misstatement. We intended to convey that each axis has a **baseline with minimal vibration**. Differences in baseline values may be due to **confidential adjustments made to the raw data** by the instructor.
3. **Q:** On page 19, how does the boxplot show the information? Why not use Q1, Q2, Q3, but focus on outliers?
   A: Q1, Q2, Q3 mostly contain **stable (non-vibrating) data**, so they are less informative. Outliers highlight **periods of vibration**, which are more relevant for analysis.
4. **Q:** Why place sensors in two locations despite high cost?
   A: Certain features (e.g., Xb0) may only be captured effectively along **specific axes**, which can be verified through **dimensionality reduction results**.
5. **Q:** Gantt chart needs to be included in the written report (week 18 section).
   A: Acknowledged and added.
6. **Q:** How should the PM handle situations when team coordination is difficult?
   A: Work **cooperatively**; all team members are willing to help each other.
7. **Q:** Why use boxplots? Were no statistical tests performed?
   A: Concerns that **X, Y, Z axes are correlated** could invalidate many standard tests.
8. **Q:** Are there statistical tests that can confirm significant differences under different loads?
   A: Yes, these were **presented in the second report**.

(II) May 7, 2024 – Second Report Q&A

Instructor Questions & Team Responses

1. **Q:** Is the moving window method used to observe data over a certain period? How is the window size determined?
   A: Window size of **two points** was used, based on references and common practice in the literature.
2. **Q:** There are two preprocessing methods. Which one is used and why?
   A: Chose **extracting the full waveform head and tail**, because **threshold-based methods may accidentally remove high-frequency waveforms**, even

when their variance is small.

3. **Q:** How does the literature data differ from ours? Will their grouping method differ as well?
   A: Baselines differ across forces. For structures that are very similar, **entropy is ineffective**, so this approach was eventually discarded.
4. **Q:** On page 21, after preprocessing, are there any remaining unprocessed data? Any "small blind spots"?
   A: Due to the large data volume, some **minor segmentation inaccuracies are unavoidable**. This is a necessary compromise to **automate preprocessing across the entire dataset**.
5. **Q:** After preprocessing, should entropy and Fourier features be applied to the **same classifier**? Should features be tested before modeling?
   A: Since entropy-based features performed poorly in preliminary tests, the **entropy-based approach was removed**. Only Fourier-based features were used moving forward.

(III) June 4, 2024 – Third Report Q&A

Instructor Questions & Team Responses

1. **Q:** Much of the presentation focused on feature extraction. Are the sources of reference existing methods, or custom?
   A: Features were extracted by applying **Fourier transformations** to the data, using standard modules for the required features.
2. **Q:** Did you consider variables that only emerge when **X, Y, Z axes are combined**? Are they unimportant?
   A: Combining X, Y, Z axes may lead to **cancellation of signals**. During the first report, statistical tests on combined axes were **not effective**. The instructor emphasized that the **key point is identifying which features are important**, rather than forcing combination.
3. **Q:** The model seems fixed. Have you considered **when to update the model**? How can customers trust the health index? What about after-sales service?
   A: Accuracy metrics will be provided to build **customer confidence**. If accuracy falls below a threshold, the system will **notify relevant personnel** that the model may need updating.
4. **Q:** Since the data is based on **spectrograms**, is there literature supporting the use of the difference between healthy and abnormal spectra to classify anomalies? Could establishing a "normal spectrum" help generalize across different robotic arms?
   A: Literature is limited because **robotic arm types vary**, and differences

between brands are unclear. Previous references using different arms (e.g., entropy-based methods) were **not very effective**.

13 References

1. Evaluation of Entropy Analysis as a Fault-Related Feature for Detecting Faults in Induction Motors and Their Kinematic Chain (https://www.mdpi.com/2079-9292/13/8/1524)