

P002: IoT-based Virtual Health Determination

巨量資料專案書面

姓名學號 410978004 統計四 郭依璇
410978038 統計四 戴衣伶
410978051 統計四 林瑋珈
410978055 統計四 薛珮妤
410978058 統計四 李博業
411078018 統計三 侯言蓁
411078038 統計三 陳皓鈞

目錄

一、前言	3
二、專案背景說明與目的	3
(一) 專案背景	3
(二) 專案目的	3
三、專案執行計畫	3
四、資料觀察	4
(一) 資料說明	4
(二) 資料觀察	5
五、資料前處理	7
(一) 確認是否有 degradation	8
(二) 移除非振動期訊號資料	9
六、分析方法原理介紹	13
(一) 熵	13
(二) 快速傅立葉轉換	13
(三) 分段擷取最大值	13
七、分析與驗證結果	13
(一) 熵	13
(二) 快速傅立葉轉換	14
(三) 分段擷取最大值	17
八、輔助決策工具設計稿	19
九、結論	20
十、組員分工名單	22
十一、組員對專案貢獻	22
十二、會議記錄	24
(一) 每週會議紀錄	24
(二) 三次報告 Q&A	27

一、前言

在現代製造業中，機械手臂扮演著至關重要的角色，然而其傳動系統若出現問題，將直接影響生產效率和產品品質。目前的檢查程序需要停機並耗費大量人力，且難以預防突發故障。因此，本專案旨在建立一套基於振動感測器數據的機械手臂健康監測方案，以提高生產效率並降低設備故障對產能造成的損失。透過機器學習方法，我們將探討振動與傳動健康度的關聯，以建立預防性保養機制，並為製造業帶來更高效、更可靠的生產環境。

二、專案背景說明與目的

第一節、專案背景

機械手臂是工業生產中常見的自動化裝置，它們通常承擔著重要的任務，如組裝、搬運、焊接等。然而，由於長期運行和磨損，機械手臂可能會出現各種問題，如軸承磨損、鬆動、失調等，這些問題會導致機械手臂的性能下降、故障甚至損壞。

而振動感知器可以用於監測機械手臂的振動情況，通過分析振動數據可以了解機械手臂的運行狀態和健康狀況。振動數據中包含了豐富的信息，如振動頻率、幅值、波形等，這些信息可以用於判斷機械手臂是否存在異常振動，從而評估其健康度。

第二節、專案目的

通過對振動感知器數據進行實時監測和分析，可以實現對機械手臂健康狀態的及時診斷和預測。當機械手臂出現異常振動時，系統可以發出警報並及時採取措施進行維修和保養，從而減少設備損壞和生產中斷的風險，提高設備的可靠性和生產效率。

三、專案執行計畫

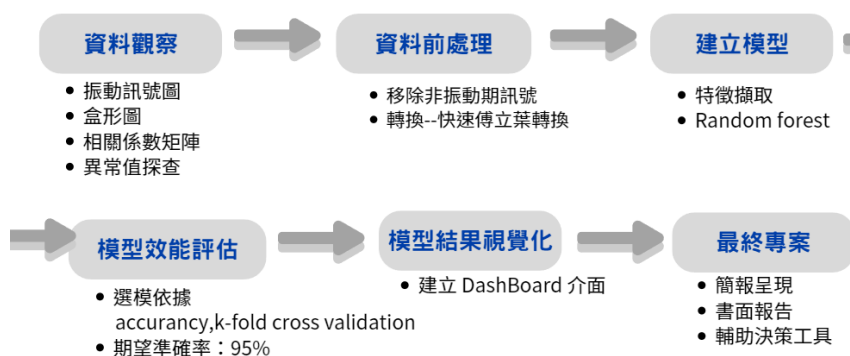


圖 1: 執行計畫

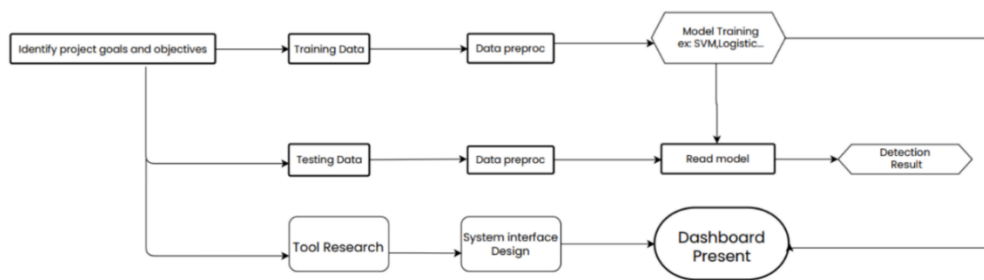


圖 2: 流程圖

進度	113 學年下學期													
	5	6	7	8	9	10	11	12	13	14	15	16	17	18
相關文獻研析														
觀察資料&資料前處理														
模型設計想法														
第一次報告														
模型測試及分析方法														
模型效能驗證和調整														
第二次報告														
驗證結果														
模型結果視覺化														
第三次報告														
書面報告														

圖 3: 甘特圖

四、資料觀察

第一節、資料說明

資料來自安裝於機械手臂的三軸加速規所感測到的 x 、 y 、 z 三軸向振動訊號。共有四顆三軸加速規，分別安裝於：

- * X_a - 位於機械手臂水平作動方向傳動軸的馬達側
- * X_b - 位於機械手臂水平作動方向傳動軸的惰輪側
- * Y_a - 位於機械手臂垂直作動方向傳動軸的馬達側
- * Y_b - 位於機械手臂垂直作動方向傳動軸的惰輪側

再對相異方向的傳動軸施加不同程度的負荷，分別如下：

- * 水平方向：65、80、95、130，其中正常負荷約於 80 左右
- * 垂直方向：220、260、300、380，其中正常負荷約於 260 左右

採集設置：

固定張力下，機械手臂以固定速度、模式作動約 5 - 6 分鐘，振動訊號每隔 5 秒儲存成單一檔案。

第二節、資料觀察

(一) 振動訊號圖--異常值

透過機械手臂上的加速規紀錄的振動訊號繪製成振動訊號圖(如下圖 4)，並觀察資料的振動情形，發現機械手臂的振動幅度會因為機械手臂的振動與否而有所改變。

當機械手臂暫時靜止時，振動訊號圖會呈現平穩狀態，約落在 0.08 至 3 的區間，此因加速度(g)的振幅會隨加速規安裝位置而有所不同；但當機械手臂運作時，其加速度振幅會呈一波峰為非振動期訊號的 2 至 3 倍之波形。因此在進行分析前，我們需先將平穩處的資料移除，即以機械手臂運行中的振動資料進行分析，避免模型受到大量靜止資料影響。

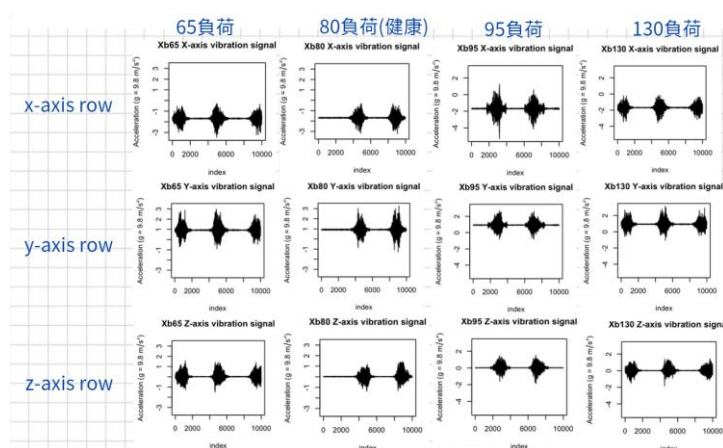


圖 4：振動訊號圖

(二) 盒形圖

從盒型圖可以觀察安裝在各個位置下的加速規，不同負荷的三軸振動均在相同基準上作動，無法以肉眼分辨，僅振動的幅度略有不同。

例如安裝在水平作動馬達測(Xa)下的加速規，負荷 65 的 X 軸振動訊號與負荷 80(健康狀態)相比有較低的加速度，Z 軸振動則是有較高的加速度；而馬達測(Xb)則是負荷 95 的 X 軸的振動訊號加速度全距較長。

此外，在垂直作動下，安裝在馬達側(Ya)的加速規在負荷 220 時，X 軸振動訊號與負荷 260(健康狀態)相比有較低的加速度，Y 軸則是負荷 220 與 300 的振動訊號加速度全距較負荷 260 短，Z 軸則是不健康狀態下的三個振動訊號加速度全距都較短(如下圖 5)；而惰輪側(Yb)的加速規在負荷 220 時，X 軸振動訊號與負荷 260(健康狀態)相比有較低的加速度，Y、Z 軸則是負荷 300 與 380 的

振動訊號加速度全距較負荷 260 長(如下圖 6)。

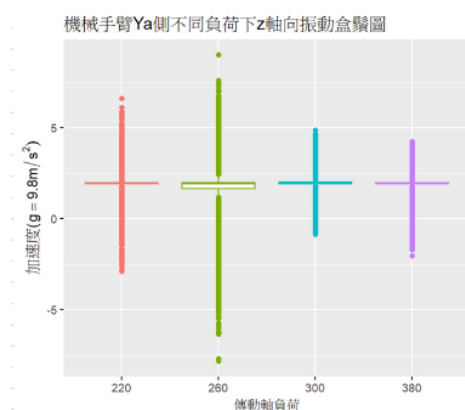


圖 5

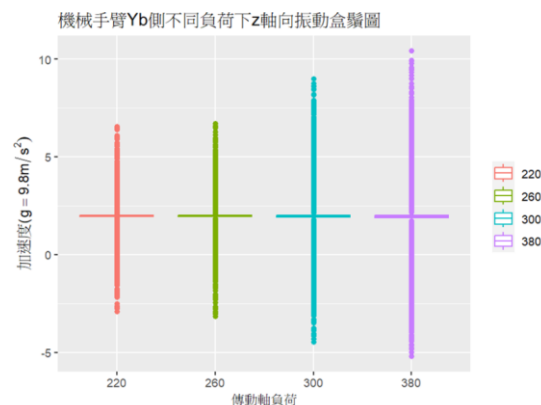


圖 6

(三) 相關係數圖

若觀察 X_a 、 X_b 、 Y_a 、 Y_b 四顆加速規在所有負荷力下， X 、 Y 、 Z 三軸間彼此的相關性並沒有太大的變化，在 X_a 以 X 、 Z 軸間有弱正相關， X_b 三軸接近無相關(如下圖 7)，垂直作動 Y_a 、 Y_b 則以 X 、 Y 軸間有些許正相關(如下圖 8)。

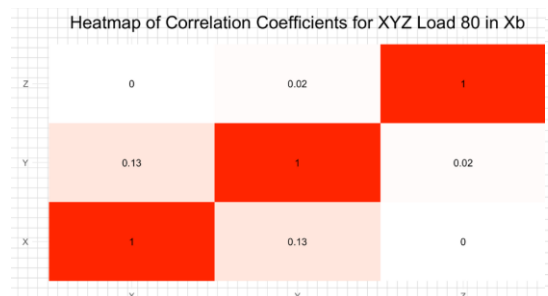


圖 7

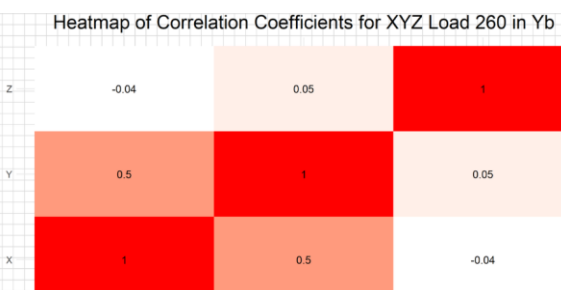


圖 8

(四) 資料視覺化

在做資料探索時，我們發現垂直作動在不同施壓力下的波型是比水平作動來的更加明顯的，像是馬達側負載力 260 相較 220 就更密集一些(如下圖 9)，而惰輪側負載力 380 相較 220 更密集一些(如下圖 10)，所以一開始猜測垂直作動的分類效果可能會較水平作動好。

垂直作動方向不同負載力的資料時序圖-馬達側Ya (x軸單位:筆數;y軸單位:加速度($g = 9.8m/s^2$))

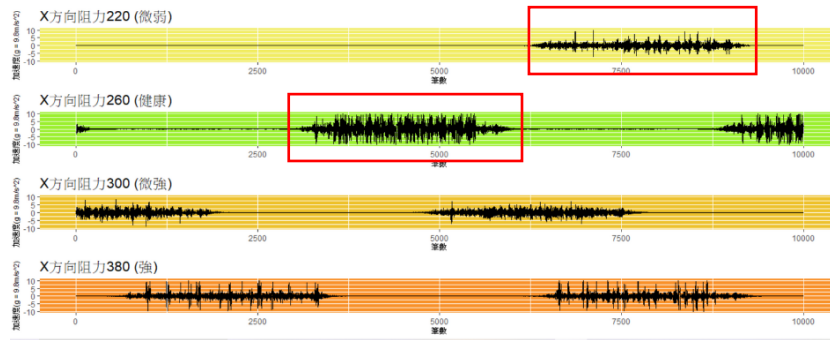


圖 9：Ya 資料時序圖

垂直作動方向不同負載力的資料時序圖-齒輪側Yb (x軸單位:筆數;y軸單位:加速度($g = 9.8m/s^2$))

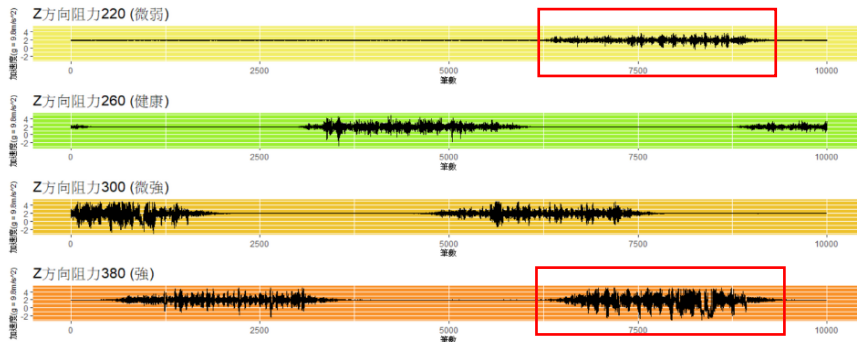


圖 10：Yb 資料時序圖

五、資料前處理

透過文獻探討，我們得知了機械手臂是會漸漸損耗，從而導致它產出的振動訊號會有誤差，也就是 degradation。並且，在資料觀察中有提到，機械手臂在非作動期間所產生的資料並不在我們想討論的範圍內。

為了減少後續診斷結果可能造成的誤差，所以在前處理階段想確認是否有 degradation 與將平穩資料移除。

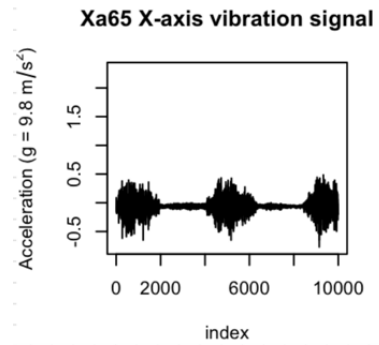


圖 11：原始資料示意圖

第一節、確認是否有 degradation

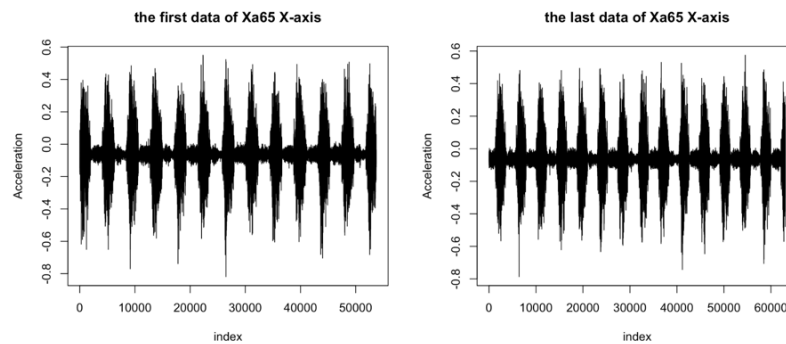


圖 12 最前面與最後面的資料的振動訊號圖

將 Xa65 的資料分成前中後段做盒型圖，示意圖如下圖 13

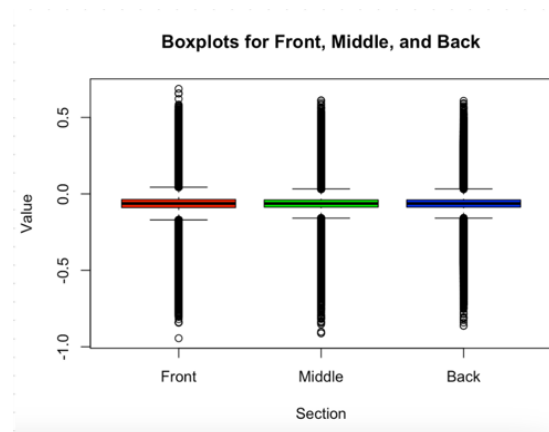


圖 13

根據以上兩張圖片，可以看出沒有肉眼可見的 degradation，故判定沒有 degradation。並且資料背景確實也是機械手臂振動五分鐘的訊號，所以判定此

資料沒有 degradation 是合理的。

第二節、移除非振動期訊號資料

機械手臂的每次作動會伴隨一段空白期，而此段空白期我們稱為非振動期訊號(如下圖 14 框出部分)。非振動期訊號會干擾最終此機械手臂為正常異常之判斷，於是必須移除非振動期訊號之資料。

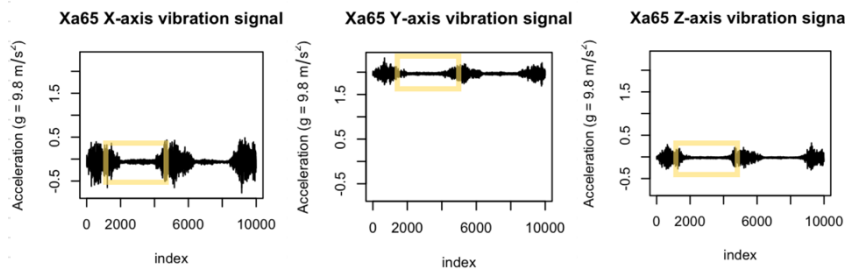


圖 14：Xa65 三軸振動訊號圖

下面我們將嘗試用三種不同方法來移除非振動期訊號，包含窗口變異數閾值及找出波的頭尾(兩種方法)，並分別討論每種方法之優劣。

(一) 窗口變異數閾值

對 X, Y, Z 三軸的絕對值進行加總並命名為 abs，並透過前三後三的移動窗口，找出該窗口內 abs 的變異數(如下圖 15)，並設一個閾值，若是變異數低於此閾值，則視為機械手臂停止時的值必須篩除。

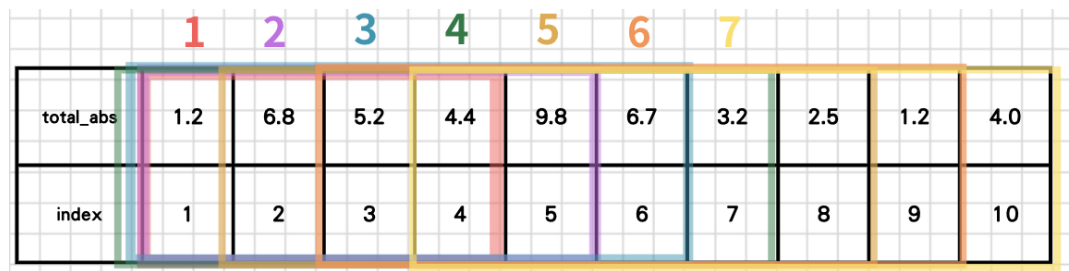


圖 15：窗口內 abs 變異數示意圖

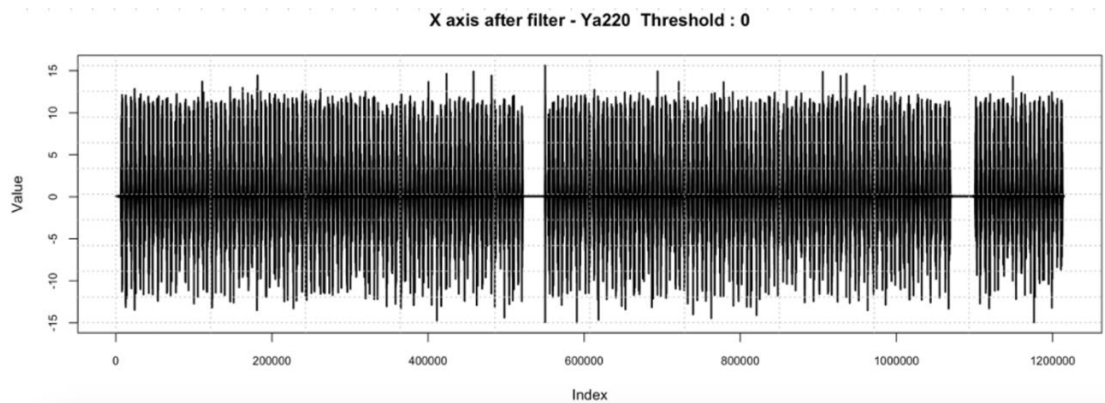


圖 16: 尚未使用變異數閾值去除平穩區段的振動訊號圖 (以 Ya220 舉例)

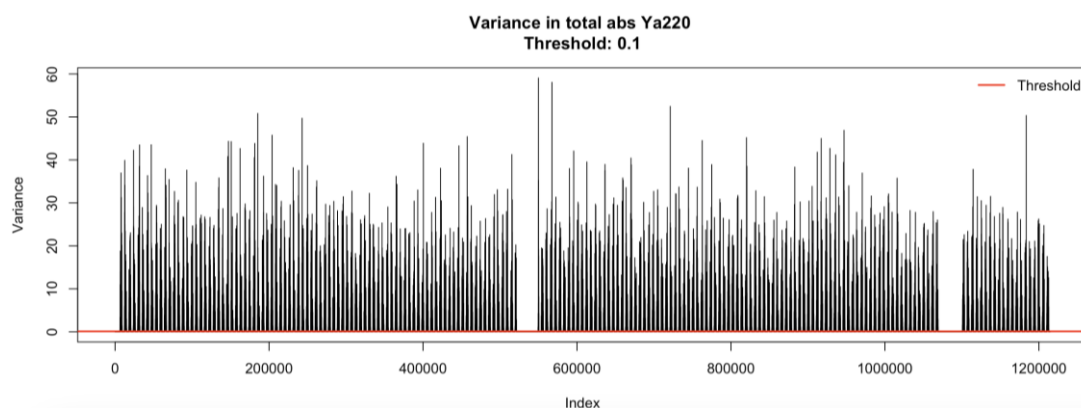


圖 17: 變異數閾值直方圖範例

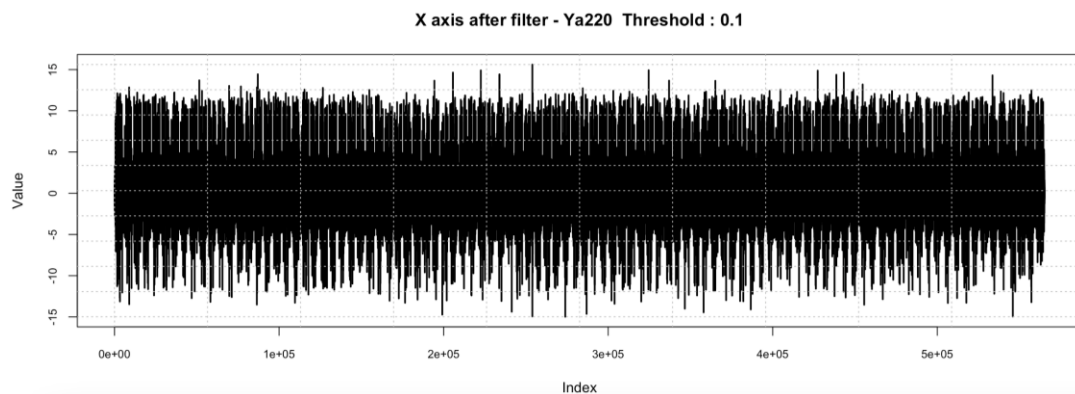


圖 18: 經變異數閾值去除平穩區段的振動訊號圖 (以 Ya220 舉例)

優點：程式執行效率高、可自動化

缺點：有可能刪到波內之值 (ex. 波內值恰好都在高峰，故幾乎無變化，此時變

異一樣會很低，有可能低於閾值而被刪掉)

(二) 找出波的頭尾--x 軸 y 軸 z 軸 每一軸皆找出各自平穩區段的 range，再找出頭尾。

表 1: Xa, Xb 各負荷下的各軸之非振動期訊號的 range

	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

表 2: Ya, Yb 各位置各負荷下的各軸之非振動期訊號的 range

	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	-2~2	-2~2
Yb\$Y	-1~1	-1~1	-1~1	-1~1
Yb\$Z	0.5~4	0.5~4	0.5~4	0.5~4

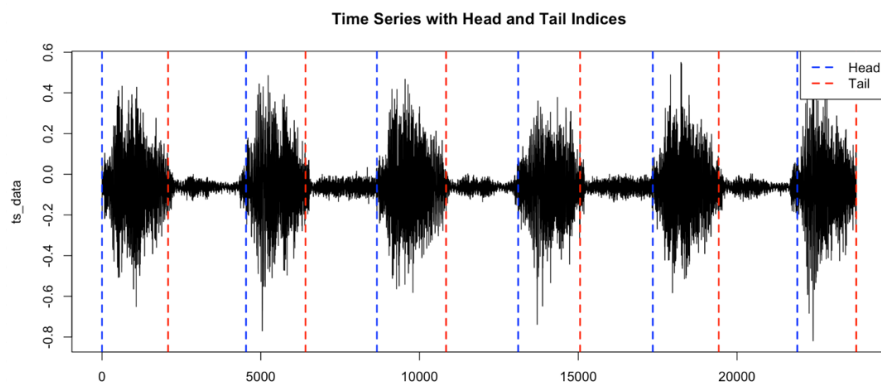


圖 19: 精確切分出波的頭尾之圖

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

圖 20:方法(二)匯出之資料長相

優點：可確保波內之值不會被篩除、幫助後續傅立葉轉換、可以根據各軸找出較為精確的振動期訊號波。

缺點：因波長不一致導致自動化較沒效率，需找出各位置各負荷下的各軸之非振動期訊號的 range。

(三) 找出波的頭尾—根據 x 軸找出 x 軸非振動期區段的 range，並且找出波的頭尾，y 軸與 z 軸沿用 x 軸頭尾的 index 來找出頭尾。

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	

圖 21:方法(三)匯出之資料長相

優點：可確保波內之值不會被篩除、幫助後續傅立葉轉換、波長一致、只需要單軸的非振動期區段的 range

缺點：自動化較沒效率，需找出各位置各負荷下的非振動期訊號的 range

六、分析方法原理介紹

第一節、熵

根據由 Y. Jaen-Cuellar 等人所撰寫之參考文獻 Evaluation of Entropy Analysis(詳見第 28 頁)，主題亦為故障分析，作者由熵的角度切入分辨機器是否故障，因他們認為用熵擷取波的訊息較有效率，且能量化信號的不確定性及多樣程度。

其中文獻使用六種熵，分別為 Spectral Entropy (譜熵)根據傅立葉轉換後的各頻率能量計算、Renyi Entropy(瑞麗熵)引入 Alpha 參數以調整機率較高與較低處的權重、Permutation Entropy(排列熵)透過子序列的排列組合計算複雜程度。而 Approximation Entropy(估計熵)、Sample Entropy(樣本熵)及 Fuzzy Entropy(模糊熵)則引入移動窗口的概念計算波的相似程度。

第二節、快速傅立葉轉換

傅立葉轉換是一種能將時域資料轉換成頻域資料的方法，被廣泛應用於分析訊號資料和訊號偵測，作法是將訊號波拆解成許多不同頻率的正弦波和餘弦波的組合波，並記錄這些組合波於原波形中的振幅，建立頻譜圖後就能觀察到原波形的許多細緻特徵，基於資料性質可分為連續傅立葉轉換及離散傅立葉轉換，此次研究適用於離散傅立葉轉換，因此選擇由離散傅立葉轉換衍伸的快速傅立葉轉換作為資料前處理的方式，有效減少運算資源。

第三節、分段擷取最大值

因原特徵擷取方法有的準確率仍有上升空間，所以思考是否有其他更好的特徵擷取手法，由於頻譜圖每一區段的變異程度不盡相同，因此產生分段擷取特徵的想法。又因振幅極大值是較原始的頻譜特徵不用做其他參數調整或數值轉換，且參考傅立葉的特徵重要性，四個位置最高中 Peak Amplitude 是最好擷取的，所以首先嘗試分段擷取極大值。

七、分析與驗證結果

第一節、熵

經過閱讀文獻並實驗後，我們發現由於此篇文獻做故障分析的機器是直流馬達，回傳的信號主要是電壓/電流的脈衝信號，其中速度是重要的因子，反觀本研究，對於機械手臂而言，精準控制負載力才是重要的，且其主要依靠感知器回傳信號。

另外，實驗設計上的方式也不同，例如這篇文獻的變因有直流馬達運行的頻率及齒輪箱的速度，但本研究只有針對機械手臂的負載力，所以由熵擷取特徵的效果最終並不如預期。以下將說明在分析中相關的假設檢定(包含 Kruskal - Wallis test 與 Fisher Discriminant Score)。

Kruskal - Wallis test 為一無母數檢定，透過比較排名中位數檢定四組資料是否來自同一母體，其中需滿足四組資料樣本間獨立的要求，但不需來自常態分佈。而在本分析資料的檢定結果顯示，幾乎所有的特徵都拒絕不同負荷的資料來自同一母體。

Fisher Discriminant Score 用於評估特徵在區分不同類別樣本時的有效性，透過計算組內與組間的散佈矩陣，衡量特徵判別不同組別的能力，通常 FDS 分數以 1 為界線，大於 1 代表該特徵有能力區分不同組別。在本分析資料中，不同位置的檢定結果都不相同，其中 Xa 與 Yb 的檢定結果較差，僅少部分變數的 FDS 分數大於 1，而 Xb 與 Ya 則有較多變數的 FDS 分數大於 1，表示這些變數區分不同健康程度的能力較佳。

第二節、傅立葉轉換

此次研究的特徵擷取方式使用快速傅立葉轉換，將已分段的各個波的 x, y, z 軸振動從時域資料轉為由正弦波及餘弦波組成的頻域資料，並從頻譜上存取下列波形基本特徵：

1. 變異數、分位數、平均數、中位數：整段波的敘述性統計。
2. 主頻率：頻譜圖中振幅最大的頻率。
3. 峰值數量：計算每個頻率的振幅和鄰近的振幅差異超過振幅最高減最低的一半，即定義為峰值，並記錄其數量。
4. 峰值振幅：所有峰值中的最大振幅。
5. 頻譜能量：對所有振幅做平方相加，可看作整段波的能量總和。
6. 頻譜熵：頻譜的複雜度與不規則性，值的高低可以大致看出頻率的分佈情況。
7. 頻率帶寬：根據頻譜能量，取出包含主要能量的頻率範圍。
8. 頻率中心：訊號能量分佈的重心。
9. 頻譜尖銳度：測量頻譜形狀的峰度，反應出頻率成分的集中度。

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

圖 22: 特徵擷取後的資料表樣式

擷取完以上特徵後，因為 x、y、z 軸都分別包含以上這所有的特徵，為了減少模型複雜度和增加模型解釋性，故我們使用費雪特徵檢定來篩選各個位置的顯著特徵。

特徵	Xa	Xb	Ya	Yb
變異數		√	√	
分位數		√	√	
平均數		√	√	
中位數		√	√	
主頻率			√	
峰值數量			√	
峰值振幅		√	√	
頻譜能量		√	√	
頻譜熵		√	√	
頻率帶寬		√	√	√
頻率中心		√	√	
頻譜尖銳度		√	√	√

圖 23: 費雪特徵檢定的顯著性

可以看到在 Xa 和 Yb 的特徵對區分健康度的能力較弱，而 Xb 和 Ya 的特徵的能力相對好很多，為了避免後續模型訓練會產生太大的差異，因此統一保留所有特徵進入模型中。

模型選擇：使用隨機森林模型(Random forest)

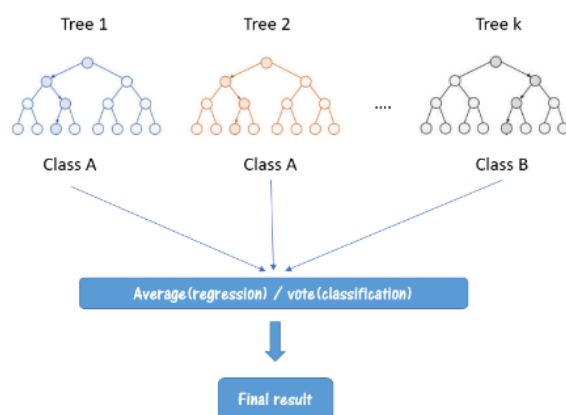


圖 24:隨機森林結構

考慮到特徵變數很多，不確定不同軸之間是否存在交互作用，使用隨機森林可以有效率的試驗 x、y、z 軸的交互作用，並且因為每個決策樹使用的特徵變數都是隨機抽取的，可以增加模型的多樣性，減少發生過度擬合以及共線性的影響。

位置	模型準確率
水平作動馬達側(Xa)	0.8813
水平作動惰輪側(Xb)	0.9281
垂直作動馬達側(Ya)	0.9918
垂直作動惰輪側(Yb)	0.9836

圖 25:模型結果

隨機森林的分類結果，可以看到除了 Xa 位置以外，其餘都有達到九成以上的準確率，而從受力方向來說，垂直作動的分類效果明顯較水平作動好。

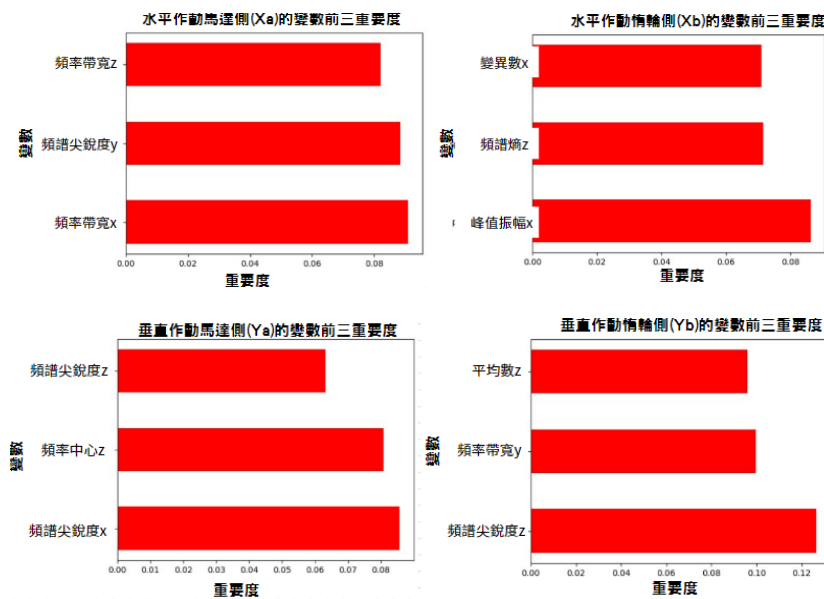


圖 26: 各位置的隨機森林前三重要度特徵

從上圖之特徵重要度來看，Xa 位置最重要的變數為 x 軸的頻率帶寬，Xb 位置為 x 軸的峰值振幅，Ya 位置為 x 軸的頻譜尖銳度，而 Yb 位置為 z 軸的頻譜尖銳度。其中 Yb 位置的前三變數重要度比例較另外三個位置高，雖然看不出有什麼特別的規律，但在後續其他特徵抽取上還是能提供資訊，並根據不同位置

來調整特徵的選擇。

第三節、分段擷取最大值

以下為分段擷取最大值示意圖(圖 27)，以切片數 9 為例，紅點為每一區段的極大值。

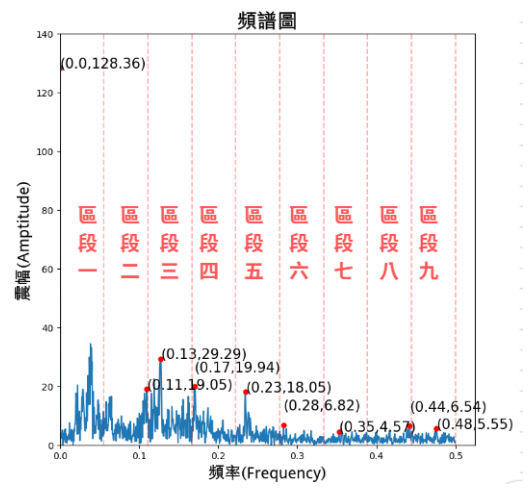


圖 27:分段擷取極大值示意圖

以切片數 9 為例，X, Y, Z 軸會各有其區段極大值，如下圖 28 所示。

此波羅 Index	此波羅 Index	X軸-區段一	Y軸-區段一	Z軸-區段一	擷取 力程 標	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	4.509690	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	5.399884	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	1.950968
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

圖 28：資料表範例

然而，我們發現每一位位置適合的切片數可能不同，所以嘗試每個位置先選出一組固定的訓練集及測試集，並對不同切片數做測試觀察效果，下圖 29 有四個位置的準確率折線圖。

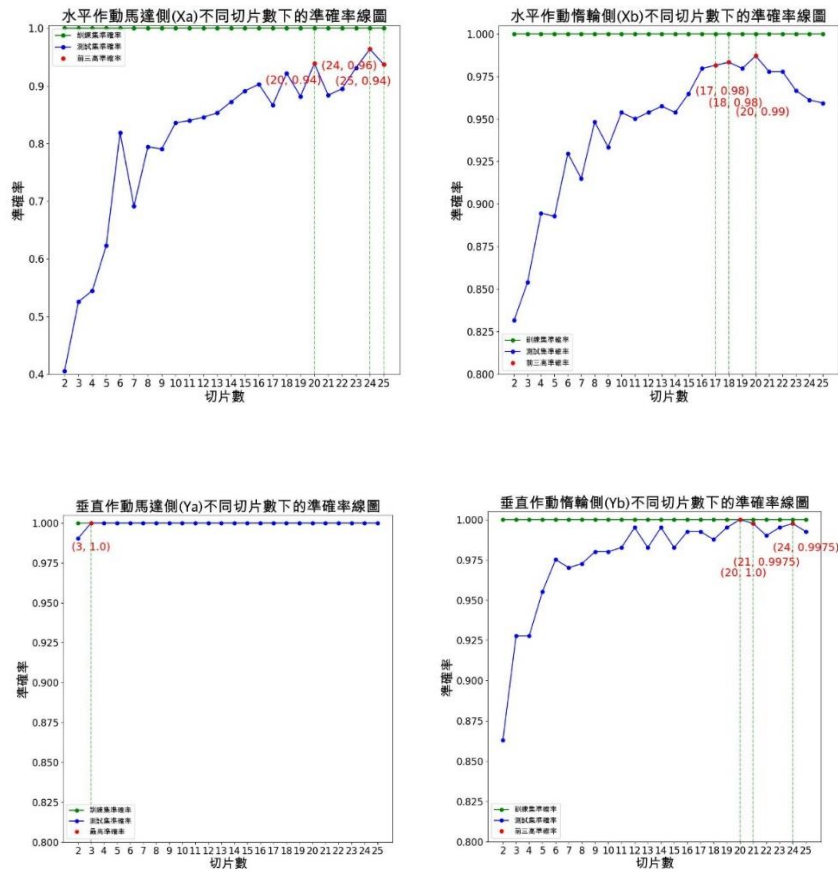


圖 29：四位置準確度折線圖

其中標出紅點的為我們認為各位置表現較良好的切片數，再透過交叉驗證選擇準確率較穩定的切片數，如下圖 30 所示。

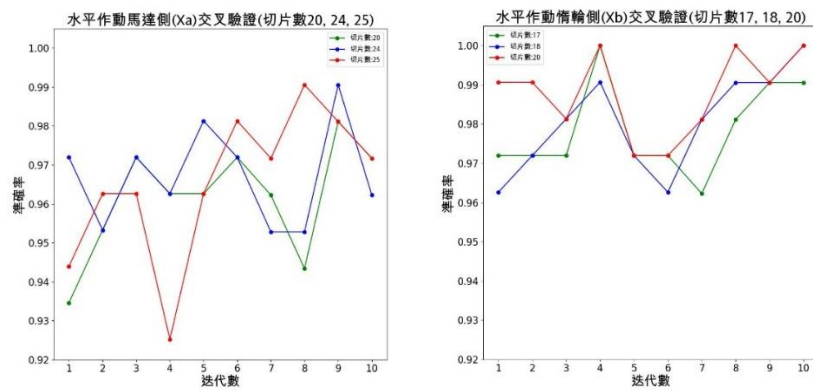


圖 30

水平作動馬達側選擇切片數 24，惰輪側選擇切片數 20，因為表現較穩定且高。此外，若要檢測水平作動負載異常，安裝感知器於惰輪側是比馬達測更好

的，因為準確率較高。

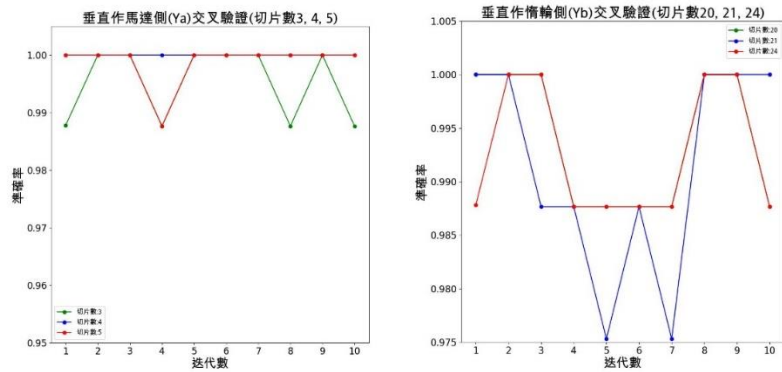


圖 31

垂直作動馬達側選擇切片數 4，惰輪側切片數 24 與 20 結果相近，為了減少計算量選擇切片數 20。此外，若要檢測垂直作動負載異常，安裝感知器於馬達側是比惰輪側更好的，因準確率較高且穩定。

八、輔助決策工具設計稿

設計稿的預想建立在我們假設使用此輔助工具者皆為機械手臂振動資料的持有者。進入頁面後，使用者須先選擇其欲觀測的作動位置，並控制上傳文件數量，由於 R shiny 執行的內存空間限制，我們將上傳文件數量上限設定在 25。點擊處理文件後，使用者上傳之資料將進入 server 端進行資料前處理、快速傅立葉轉換、特徵擷取及模型檢驗。

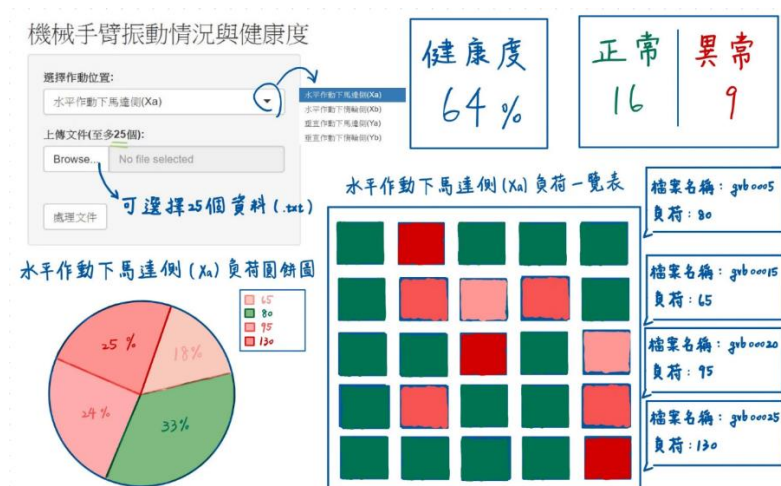


圖 32：輔助工具界面之示意圖

根據初始選擇之作動位置，輔助工具左下方預計顯示出在其作動位置下各負荷之占比的圓餅圖，其中健康負荷呈現綠色，其餘則為紅色（暫定）；右上方將顯示上傳資料中健康程度占比以及健康、不健康之各別數量；此外，為了讓使用者能夠更準確的了解上傳文件中各別的健康度，我們希望在右下方設計為健康度一覽表，當使用者將滑鼠游標移至每一方格上，皆能顯示出其檔案名稱及其負荷；在上方空白處則預計放入下次更新模型之時間以及客服專線電話、相關人員信箱等。

另外，站在使用者角度思考，我們認為並非所有使用者對模型建立及統計都有基礎概念，其主要想了解的只有上傳資料之健康狀況，因此我們決定不顯示出資料的原始數據及擷取完波形頭尾之振動訊號圖，僅呈現簡單明瞭的健康度、手臂良好與否等資訊。

補充說明：

1. 若是使用者選擇作動位置與實際上傳文件的作動位置不一致，會跳出彈窗顯示錯誤。
2. 此 dashboard 功能可以重複執行，不需要進行重新整理。

為了讓使用者相信模型的準確率，我們提供的基礎售後服務為：

1. 準確度保證：當準確度低於 95% 時，將會寄送警示信件給公司，公司將安排技術人員前往更新維修。
2. 客製化模型：利用客戶資料不斷訓練模型，以提供更高的準確率，使客戶相信輔助決策工具之結果。

若進一步購買 VIP 方案，使用者將享有兩個月定期檢查更新模型，且提供健康檢測報告以及優先疑難排解之權利。

更多 dashboard 實作與操作細節請參考以下網址：

https://github.com/Iris910531/BigData_course

九、結論

在前面分析與驗證結果中提到了的三種作法(熵、快速傅立葉、分段擷取最大值)，其中熵因假設檢定結果不佳，故最終我們並未採用此方法。而現在有兩種特徵擷取方式可選擇，我們期望結果是穩定且準確率高的，所以分別對各位置做兩種方法的交叉驗證比較，下圖是結果。

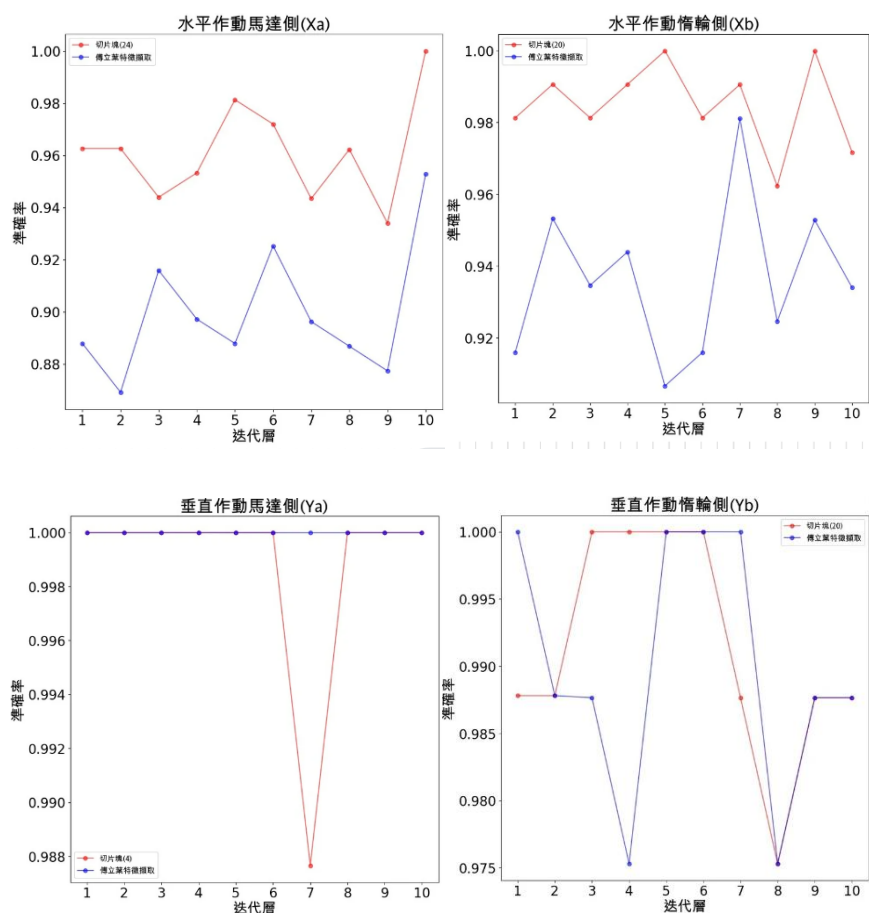


圖 33

圖中我們可以發現分段擷取極大值普遍表現是較好的，後續為了確保所有預測方法的一致性，故最終我們選擇使用分段擷取極大值的方法。

根據前述分段擷取極大值個切片數下的準確率結果，可以看出 Xb 在各切片數下比起 Xa 準確率都較高且較穩定，因此我們認為水平作動適合在惰輪側放感知器做異常偵測。同理，因 Ya 準確率表現比 Yb 好，故垂直作動則適合在馬達側放感知器做異常偵測。

我們認為分段擷取振幅極大值所訓練出的模型效能高於所有特徵的訓練模型，其主因為分段擷取可以將每一波切得更細緻，讓模型可以更精準的訓練。未來我們希望可以根據前述的費雪檢定將較重要的變數以分段擷取的方式進行模型訓練，希望可以得到更穩定、更高效能的模型。

對於此次專案，我們認為若是能夠提供更多專案背景介紹，例如實驗影片或流程圖，可以找尋更適當的文獻當作參考。另外，對於實驗我們認為廠商可以增加更多種負荷力，讓模型更豐富且更能廣泛應用。

十、組員分工名單

PM	DS	SD
<ul style="list-style-type: none"> • 侯言蓁 <ul style="list-style-type: none"> ◦ 甘特圖+簡報美編+會議記錄+期末書面製作 • 戴衣伶 <ul style="list-style-type: none"> ◦ 簡報美編+流程圖製作 	<ul style="list-style-type: none"> • 林瑋珈 <ul style="list-style-type: none"> ◦ 傅立葉轉換+特徵擷取 • 郭依璇 <ul style="list-style-type: none"> ◦ 資料前處理 • 薛珮妤 <ul style="list-style-type: none"> ◦ 熵的統計檢定+文獻探討 • 李博業 <ul style="list-style-type: none"> ◦ 熵的計算+文獻探討+訓練模型 • 陳皓鈞 <ul style="list-style-type: none"> ◦ 傅立葉轉換文獻探討+訓練模型 	<ul style="list-style-type: none"> • 戴衣伶 Dashboard 設計&實作 • 郭依璇 <ul style="list-style-type: none"> ◦ Dashboard 設計&實作+github處理 • 陳皓鈞 Dashboard 實作 • 薛珮妤 Dashboard 實作

圖 34：組員分工名單

十一、組員對專案貢獻

郭依璇：在本專案中，我負責振動訊號前處理的自動化，確保有效地去除雜訊，並提升資料的準確性和可靠性。此外，我組織並領導了多次專案會議，確保團隊的協同作業與順利推進。在會議中，我亦負責部分的會議紀錄，記錄關鍵決策和行動項目。在儀表板設計與實作方面，我主導了整個設計過程，並負責實際的編程實作，將前處理數據、模型應用及最終結果集成並展示在儀表板上。此外，我負責整合所有的程式碼，包括資料前處理、模型應用和結果呈現，確保專案的整體一致性與高效性。最後，我將完整的儀表板實作程式碼與相關環境和套件相容性上傳至 GitHub，方便其他團隊成員及未來的使用者取用，確保專案的可持續性與可擴展性。

戴衣伶：在這次專案中，我主要負責簡報製作及美編、專案執行計畫及流程圖設計，以及透過 R Shiny 設計和部分實作 Dashboard。在簡報方面，我精心設計和製作專業簡報，確保關鍵信息的清晰傳達並提升視覺吸引力。在專案計畫和流程圖設計中，我詳細規劃及描述專案的各階段和步驟，為專案的成功執行提供堅實的基礎，確保工作有序高效進行。此外，我通過 R Shiny 設計並參與了部分實作 Dashboard，實現關鍵數據的即時可視化呈現，確保數據的準確性和及時性為決策提供了強有力的支持。我的工作提升了專案的效率和效果，為專案的成功奠定了堅實基礎。

林瑋珈：在這個專案中，首先，我先深入了解快速傅立葉變換（FFT）的原理及其廣泛的應用情境，確保理論基礎扎實且能夠靈活運用。其次，我將前處理後的資料轉換為頻域資料，為後續的分析和模型訓練奠定了基礎。再來，針對機械手臂的應用，我多方查找相關文獻，萃取了大量有價值的特徵，這些特徵對訓練高效模型至關重要。最後，我完成所有數據的特徵擷取工作，為專案的成

功做出了重要貢獻，使我們能夠在訓練模型時得到高準確率。此外，我的報告、撰寫能力在簡報與書面報告中發揮極大的作用，進一步提升了專案的整體品質。

薛珮妤：在本專案中進行了資料視覺化工作，利用盒型圖展示資料分佈，從中發掘資料前處理條件，提升資料截取的準確性；接著在模型構建前，進行相關文獻與傅立葉轉換研究，以科學嚴謹的統計檢定方法篩選出最具分類和預測價值的特徵；最終，我以 R Shiny 製作了互動式的儀表板（Dashboard），實現了數據的動態展示和分析，提升了專案的專業性和實用性，透過互動式儀表板，使用者能夠便捷地探索和分析數據，從而提升了專案的應用價值和用戶體驗。

李博業：在這次專案中，我涉獵了資料視覺化、資料前處理、特徵擷取及模型訓練。首先，我在資料視覺化過程中觀察了各位置的不同負載力振動差異。其次，為了移除待機波段，我使用移動窗口計算變異數，並篩除低變異部分。接著，根據文獻資料，進行了六種熵的特徵擷取。然後，對快速傅立葉轉換後的頻譜圖進行切段、擷取極大值，並選擇各位置最適合的切片數。最後，利用隨機森林模型進行訓練，並通過 K-fold 交叉驗證來確保模型的穩定性，最終模型的準確率超過了 95%。

侯言蓁：在這個專案中，我詳細記錄了每周的會議紀錄和部分需要向教授請教的問題，確保能及時獲得指導和解答。首次上台報告中，我講解了專案的總體方向和目標，介紹了初步資料觀察的結果，為後續的分析和報告奠定基礎。另外我負責製作甘特圖來規劃專案的各項任務及其時間安排，確保專案按計劃進行。在三次報告中，我負責 PPT 的美編和排版工作，確保報告內容清晰、美觀，提升了報告的視覺和表達效果。我還整合了成員們分散的結果內容，負責書面報告的最終製作，包括內容的編寫、排版和校對，確保報告的完整性和專業性。通過這些工作，我展現了出色的專案管理能力和團隊合作精神，確保了專案的順利進行並大大提升了報告的品質和效果。

陳皓鈞：閱讀振動資料分析的文獻，並深入研究傅立葉轉換的應用及其局限性是我們專案的重要一環。在隊友成功擷取振動波形的基本特徵後，建立了一個包含全部特徵的隨機森林模型。這個模型不僅為我們提供了更深入的數據理解，還作為後續模型比較的重要參考。在儀表板實作的過程中，我與隊友緊密合作，並持續提供心靈層面的支持，以確保視覺化介面的結果完整呈現此次研究。

十二、會議記錄

第一節、每週會議紀錄

03/30

討論文獻: phm2012 可行性

包括特徵工程 (偏態峰態....) RUL 剩餘壽命 等概念 和 PCA(太耗時?) & FFT
先著手資料探勘、資料縮減

工作分配:維度縮減:郭、EDA:業&郭、甘特圖:侯

04/05

原思考方向錯誤:資料的收集回傳需要時間-> txt 裡面的資料間隔可能不是均勻分配&加速度較平穩處是機器靜止狀態非 pattern 的模式特徵

要把機器靜止處的資料刪掉

1. 把每列(含自己)的前後三筆作變異數,把變異很小(<0.01?)的地方視為機器靜止刪除

but 只能單看 x 單看 y 單看 z 這樣可能會有 x 被刪了但 y 沒被刪之類的情況

2. 取距離的變異(歐基里德)

3. 移動平均??正負相交問題

4. 合併 x y z? 取絕對值相加在挑選閾值

閾值選定:

x y z 的閾值標準要一樣?

1. 查機械手臂加速度多少被視為是有動作

平均+-三個標準差(but 0 很多、三倍標準差會涵蓋到部分數據 不太可行)

2. 合併 x y z 取絕對值相加在挑選閾值

3. 距離的變異(歐基里德)但抓到高點的兩筆資料間差距會很小就會被刪掉?

—最後未採用閾值作法而是用抓取波的頭尾取代

—時間序列方法可能無法採用,主要考慮傅立葉

工作分配:時序圖:璇 & 業、傅立葉:陳 & 玠、視覺化:珮 & 業、ppt 編排:侯&戴

04/22

討論 ds 找的文獻

璇、珮的不太適用,業的熵感覺可行!!!

想法:stationary 可快速傅立葉 轉換完後除噪

找頭、後面就切 2 的 n 次方? 頭尾補 0

工作分配:資料前處理:璇

04/27

跟大家解釋熵的文獻和分析方法

但時間問題後續熵&傅立葉可能先擇一，傅立葉為主

傅立葉考慮特徵:頻率、震幅(能量)、向位(波峰位置)

工作分配:前處理(切窗口):璇、快速傅立葉:陳&林、熵:業&珮、改甘特圖+生流程圖:侯&戴

05/05

ds 解釋上禮拜跑出來的結果：前處理、傅立葉、熵

&大致修整第二次報告 ppt + 分配講稿

稍微討論一下邏輯斯做法

工作分配:做相關檢定:珮

05/06

繼昨天繼續修整第二次報告 ppt

&run 一次大家的稿

05/10

傅立葉轉換後的資料特徵要留哪些?頻率帶寬?

1. 隨機森林 2. SVM 3. 邏輯斯 三種模型要跑，代入資料還要調參數

dashboard:要怎麼去掉客戶的資料的平穩段?設死範圍?能自動化的話更好

工作分配:測試及處理:璇、特徵處理:林、特徵檢定:珮、特徵視覺化:業

05/18

xa 分類效果不太好(但老師說是第二好 QAQ)

xb 大致都能區分 問題不大

模型上的問題:

xyz 分開(要不要把沒用地丟掉)或合起來?

1. 隨機森林 2. SVM 3. 邏輯斯

找變數前用平方算距離或絕對值把 xyz 合成一個變數?

dashboard:PM 要當刁鑽的客戶發問並記錄在書面裡

—老師建議可以用切片去看

05/25

Dashboard 測試資料輸入

(預設輸入的資料沒問題 長度需夠長)一個 txt 大概有四個波

振動訊號要放嗎?

no 因為客戶可能沒想看/且預測客戶給的資料長度沒問題所以沒必要

振動訊號頭尾圖

no 只是徒增煩惱？客戶知道了頭尾取不好那也沒方法解決？只要有預測結果出來其實在過程不要讓客戶一直懷疑模型可能不準

yes 檢視抓頭尾抓得好不好（如果有些資料太短不適合擷取頭尾可以建議客戶提供其他資料）但也有可能只截出一個頭尾的不算是異常？

Random forest / logistic

種幾棵樹(預設 100)/調參數？

no 客戶不懂模型

—後續只採用 random forest，因為兩個模型結果如果不一樣很尷尬

所以刪掉選擇使用模型選單

06/01

以整個波去切片，切片大概都抓 2-25 片分別去看，準確率都很不錯!!!

推測馬達放垂直，惰輪放水平會較佳

傅立葉的結果跟切片的結果比較（切片好像比較好）

之前找的頭尾（開始結束）不一樣沒法處理交互作用，後來統一用 x 軸的頭尾來擷取波

找眾多變數中在 xa xb ya yb 都有明顯的特徵拉出來。未來展望：根據變數重要度，之後可以選擇其餘重要變數進行切塊，可能會有更穩定、更準確的模型
Dashboard: 取閾值（頭尾低於 500）傾向刪除，資料異常，統一用切塊的方法看產生特徵後的容量（但不好處理

設計一次只能丟幾個（25 個）檔案比較好設定

print 出訊息顯示：前處理有判斷式可以檢驗錯誤，就不會丟進後續模型，而是直接跑到最後說明資料異常（用檔名/負荷區分 25 個檔）

結果參考老師範例那樣，用正方形區塊顏色表示，和做成圓餅圖、健康度、異常率

06/03

繼續修整第三次報告 ppt

&run 一次大家的稿

06/10

dashboard 要加的內容：

1. 顯示下次更新模型時間
2. 客服專線/相關聯絡人信箱/公司地址
3. 客戶滿意度回饋

4. vip 購買方案：兩個月定期檢查更新模型且提供健康檢測報告/優先處理客戶疑問

若無購買 vip 的基礎售後：

1. 準確度低於某水準（95%）回寄 gmail 給公司，公司再派人去更新維修
2. 利用客戶資料不斷訓練模型（客製化），以提供更高的準確率，使客戶相信 dashboard 結果

第二節、三次報告 Q&A

（一）2024/04/09 第一次報告 Q&A

老師提問：

1. 為什麼 p. 11 要留白？組員回答：固定 y 軸

2. P. 34 為什麼 X Y Z 軸說都在 0 振動？

組員回答：口誤，我們想表達三軸都各自有一個基準是沒什麼振動的，至於基準為何不同，我們認為是因原始資料機密，老師有調整過資料造成

3. P. 19 盒型圖怎麼能看出來呢？為什麼不用 Q1, Q2, Q3？而是用離群值看？

組員回答：因為 Q1, Q2, Q3 幾乎都是平穩資料，這時候看意義不大，後續才會篩除非振動期資料。

4. 為什麼 sensor 那麼貴還要放兩個地方？

組員回答：Xb0 相關會不會只有一種軸才能幫我們，可看降維效果。

5. 甘特圖要補書面報告（18 週部分）

組員回答：收到，已補

6. 組員協調遇到困難時 PM 要怎麼辦？

組員回答：齊心協力囉，組員們都很願意互相幫忙！

7. 為啥用盒型圖？沒有做統計檢定嗎？

組員回答：擔心 X Y Z 軸因為有相關所以可能很多檢定不能做

8. 有沒有統計檢定可以明確的說出不同負荷下真的訊號有差？

組員回答：在第二次報告有做呈現。

（二）2024/05/07 第二次報告 Q&A

老師提問：

1. 移動窗口是觀察某一段時間的資料嗎？怎麼找的？窗口要抓幾個點？

組員回答：兩個點，因為查資料看大部分人都這樣設所以我們也這樣做。

2. 前處理有兩種方法，要用哪個為什麼？

組員回答：選用擷取完整波的頭尾，因為閾值的篩取可能會誤刪高頻率之波（其變異數也同樣小）

3. 文獻資料跟我們的有什麼差別？他們的分組跟我們的會不會有差別？

組員回答：不同力的基準不一樣，結構很像的話用熵效果確實不好，所以後來刪掉這個做法。

4. p21 資料前處理後還有沒處理到的地方？小盲腸？

組員回答：因為資料量龐大，為了使得全部資料的前處理自動化，勢必會有部分資料切的不精確，這是使自動化可以進行必要的犧牲。

5. 處理完資料後，熵跟傅立葉是要套同個模型分類器還是？建模型前對特徵先做檢定？

組員回答：因熵後續檢定出來的效果都不是很好，所以將熵的做法刪除了。

(三)2024/06/04 第三次報告 Q&A

老師提問：

1. 很大一部分在講特徵擷取，好奇參考來源是既有擷取還是？

組員回答：用轉換後資料套模組去擷取傅立葉需要的特徵。

2. 不期待有 xyz 合在一起才有的變數嗎？是不重要嗎？

組員回答：因為怕 xyz 會被抵銷，加上第一次報告的時候，檢定做出來的效果不太好。

如果你們做出來不適合的話也沒關係，老師認為哪些特徵是重要的是關鍵！

3. 看起來模型是固定的有想過何時更新模型嗎？客戶如何相信健康度？售後服務的思考？

組員回答：補充準確率讓客戶相信準確率是高的，再安裝準確度回報。（若低於某一水準則回報給相關人員告知模型可能需要更新）

4. 因為是用頻譜圖的資料，不確定有無文獻可以佐證用健康頻譜距離異常頻譜的差距來分類異常？若是能建立正常狀態的頻譜這樣可以應對各種機械手臂資料都適用。

組員回答：文獻的部分因為機械手臂種類很多也不知道各牌之間是否有差異，因為之前參考的文獻就是不一樣的手臂然後參考效果並不是很好（熵）。

參考文獻：<https://www.mdpi.com/2079-9292/13/8/1524>