

Group Project Report

IoT-based Virtual Health Determination

廖柏竣 黃柏淞 陳威儒

June 20,2025

Contents

| | | |
|-----------|--|----------|
| 1 | 摘要 | 2 |
| 2 | 前言 | 2 |
| 3 | 研究目的 | 2 |
| 4 | 資料觀察 | 3 |
| 5 | 機器學習模型與資料前處理 | 3 |
| 5.1 | 隨機森林 (Random Forest) | 3 |
| 5.1.1 | 資料前處理與特徵提取 | 3 |
| 5.2 | 3D 卷積神經網路 (Convolutional Neural Network) | 4 |
| 5.2.1 | 資料前處理與特徵提取 | 4 |
| 5.2.2 | 模型架構 | 4 |
| 6 | 分析與驗證結果 | 5 |
| 6.1 | 隨機森林 (Random Forest) | 5 |
| 6.1.1 | 特徵篩選 | 5 |
| 6.1.2 | 模型驗證 | 5 |
| 6.2 | 3D CNN | 6 |
| 6.2.1 | 模型驗證 | 6 |
| 7 | 儀表板 | 7 |
| 8 | 結論 | 7 |
| 9 | 程式碼 | 7 |
| 10 | 參考文獻 | 8 |
| 11 | 附錄 | 8 |

1 摘要

本專案主要透過機器手臂的振動資料建立穩健且可靠的機器學習模型，以有效降低人工巡場所需的時間和人力成本，同時能夠及時精確地發現機器運作中的異常狀況。我們運用卷積神經網路（Convolutional Neural Network, CNN）和隨機森林 (Random Forest, RF) 兩種機器學習方法建立預測模型，透過多層次的數據分析提升預測準確度。實驗結果顯示模型能有效區分健康與不健康設備，且 RF 模型結果顯示 Y 軸感測器對健康狀態變化的敏感性較高，具備優先部署價值。此外，為了確保工廠人員能夠輕鬆掌握系統輸出結果，我們特別設計了直覺且易於理解的儀表板介面，將複雜的數據轉化為易讀資訊，讓現場操作人員能夠快速判讀機器狀態並做出適當的維護決策。

2 前言

機器手臂近年來廣泛應用於各領域，但傳統維護方式仍依賴人工定期巡場，往往要等到機器明顯異常才進行維修。這導致工廠面臨突發停機、生產排程混亂以及產量減少所帶來的交期壓力。在工業設備自動化與智慧製造快速發展的今日，如何有效監控設備狀態、延長使用壽命並降低非計畫性停機風險，已成為企業營運優化的關鍵課題。

隨著物聯網（Internet of Things, IoT）技術發展，我們能透過感測器收集大量資料。研究顯示，超過七成的機械故障與異常振動有關，因此振動監測技術已成為預測性維護中的核心工具，用於故障預警與健康診斷。振動監測透過感測器即時擷取設備三軸振動訊號，並藉由時域與頻域特徵分析（如 RMS、頻率中心、峭度、功率譜密度等）識別可能的磨損、鬆動或結構損傷。本研究基於此原理，建構虛擬健康判定系統，整合資料前處理、特徵工程與監督式機器學習模型，進行異常偵測與分類。

3 研究目的

本專案透過振動資料建立機器學習模型，目標為解決三大痛點：

1. 預防機械臨時故障: 機器一旦突發故障，工廠需停機維修，嚴重影響生產效率。我們運用模型預測協助工廠人員及早發現潛在問題，以便提前維護。
2. 優化感測器配置: 鑑於感測器價格昂貴，我們透過模型測試與訓練結果找出最佳測量點，有效安排感測器位置，降低不必要的數據收集。

3. 簡化數據分析: 許多感測器資料和統計數據需要專業背景才能解讀，我們設計的儀表板提供關鍵且易讀的資訊，使非專業人員也能快速掌握機器即時狀態。

4 資料觀察

資料來源為安裝於工業設備馬達側 (a) 與惰輪側 (b) 的加速度感測器所蒐集的振動訊號資料。每筆資料包含三軸方向的加速度變化，並透過不同的加壓條件模擬各種負載狀態，以區分正常與異常情況。資料具有時域的訊號特性，可進一步轉換為統計特徵及頻域特徵，為後續模型訓練提供豐富的輸入資訊。

- Xa - 水平傳動軸馬達側
- Xb - 水平傳動軸惰輪側
- Ya - 垂直傳動軸馬達側
- Yb - 垂直傳動軸惰輪側

傳動軸負荷設定 (單位保密):

- 水平方向: 65、80 (正常)、95、130
- 垂直方向: 220、260 (正常)、300、380

5 機器學習模型與資料前處理

5.1 隨機森林 (Random Forest)

5.1.1 資料前處理與特徵提取

在資料前處理階段，由於部分樣本長度明顯不足，在進行頻率與複雜度分析前，需要先行剔除以維持特徵穩定性。特徵則是將 3 軸數據平均後提取，特徵擷取設計參考多篇文獻，最終決定納入四類指標：

- 時域: RMS、Skewness、Kurtosis、Entropy、Crest Factor；
- 頻域: Freq_Center、Spectral_Kurtosis、Spectral_Entropy、RMSF；
- 時頻域: Wavelet D3 Kurtosis；
- 複雜度: Impulse Factor、Clearance Factor、Approximate Entropy。

5.2 3D 卷積神經網路 (Convolutional Neural Network)

5.2.1 資料前處理與特徵提取

在資料前處理階段，我們一樣先行剔除掉具有較多雜訊的樣本，接著，由於 CNN 以及待會要轉換的特徵特性，我們會先將資料切成相同的長度。最後將資料轉換為功率譜密度 (Power Spectral Density, PSD) 與交叉譜密度 (Cross Spectral Density, CSD)，並依據下列結構重新組合成矩陣形式：

$$\begin{bmatrix} P_{xx} & P_{xy} & P_{xz} \\ P_{xy} & P_{yy} & P_{yz} \\ P_{xz} & P_{yz} & P_{zz} \end{bmatrix}$$

其中， P_{xx} 、 P_{yy} 、 P_{zz} 分別代表 x 、 y 、 z 三個方向的 PSD，而 P_{xy} 、 P_{xz} 、 P_{yz} 則分別為對應方向的 CSD。此矩陣形式與共變異矩陣類似，並作為輸入特徵提供給 CNN 進行後續分析。

5.2.2 模型架構

在模型架構方面，確保模型在有限的樣本下能充分收斂，我們在卷積層後加入 Batch Normalization 和 dropout 以降低過擬合的風險，並透過適度最大池化 (Max Pooling) 提升運算效率並增加模型穩定性。超參數設定上，批量大小 (Batch Size) 設定為 16 且 Epochs 設定為 500。

6 分析與驗證結果

6.1 隨機森林 (Random Forest)

6.1.1 特徵篩選

本實驗以標準負荷 (80, 260) 之樣本定義為健康標籤，其餘樣本則視為異常。訓練過程採用分層 5 折交叉驗證，於每一折訓練中記錄隨機森林模型之特徵重要度 (Feature Importance)，並計算其平均值 (如圖 1 所示)。最後我們選取平均重要度大於 0.06 的關鍵特徵作為最終模型輸入，包含：RMS、Kurtosis、RMSF、Spectral_Kurtosis、Spectral_Entropy 及 Spectral_Energy。

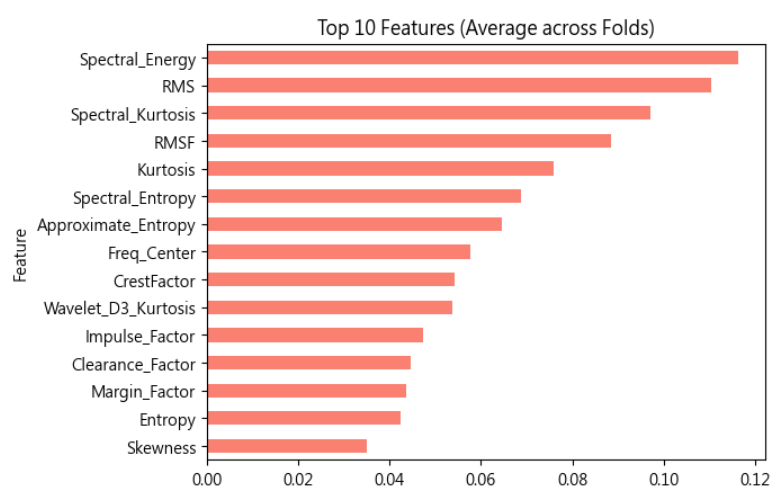


圖 1: 特徵重要度

6.1.2 模型驗證

最後我們的模型參數設定為 $n_estimators = 100$, $random_state = 42$ 且根據不同的方向 (Xa, Xb, Ya, Yb) 分別進行獨立訓練與測試，其模型評估指標結果呈現在表1。

| Direction | Accuracy | AUC | Recall | KS_Distance |
|-----------|----------|--------|--------|-------------|
| Xa | 0.8043 | 0.7532 | 0.3636 | 0.4234 |
| Xb | 0.7609 | 0.8416 | 0.1818 | 0.6000 |
| Ya | 0.9783 | 1.0000 | 0.9167 | 1.0000 |
| Yb | 0.9783 | 1.0000 | 1.0000 | 1.0000 |

表 1: 各方向模型在不同評估指標的表現 (Accuracy、AUC、Recall、KS-Distance)

由表1可以發現 Ya 與 Yb 表現穩定，AUC 接近 1，Recall 超過 0.9，KS 距離亦顯著高於其他方向，顯示其在異常判別上具高度敏感性。為了佐證個特徵在不同方向辨識能力的強弱，可繪製不同特徵下兩個群組 (健康/不健康) 在四個方向的表現，可以發現所選取的六個特徵在 X 軸方向與 Y 軸方向都可以看到明顯分離，但是在 Y 軸方向更加明顯 (詳見附錄圖4-8)。所以我們認為 Y 軸感測器對健康狀態變化的敏感性較高，具備優先部署價值。

6.2 3D CNN

6.2.1 模型驗證

在輸出設計方面與 RF 不同，本實驗我們將負荷值維持原本分類，即 X 軸方向的預測目標為 {65, 80, 95, 130}，Y 軸方向的預測目標則為 {220, 260, 300, 380}。模型訓練過程一樣採用分層 5 折交叉驗證，資料按 8 : 2 比例劃分為訓練集與測試集。CNN 模型五次交叉驗證的平均結果如圖2所示，整體而言，X 軸方向的預測準確率皆高於 0.98，Y 軸方向則高於 0.99，顯示模型在分類的表現穩定且優異。

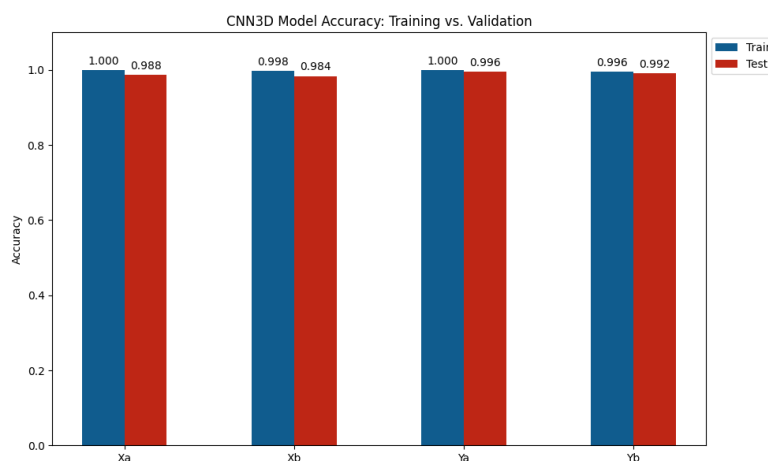


圖 2: CNN 驗證結果

7 儀表板

在模型訓練完後，為確保工廠人員能夠輕鬆掌握系統輸出結果，特別設計了直覺且易於理解的儀表板介面，將複雜的數據轉化為易讀資訊，讓現場操作人員能夠快速判讀機器狀態並做出適當的維護決策。在儀表板中只要資料輸入並選擇資料所屬的方向就可以產出 RF 或 CNN 模型的預測結果以及警示燈號 (未來實際應用上可以針對特定感測器進行串接)。



圖 3: 儀表板

8 結論

機械手臂的異常會影響精度和產能，若等到出現嚴重異常才進行停機檢測，將大幅影響生產效率。本專案致力建立預防保養機制並比較不同位置感測器資料的模型效果。我們使用了 RF 與 CNN 兩種機器學習模型，建立自動化健康監測方案來達到及早發現異常並減少損失的目標。實驗結果顯示，CNN 模型不管在哪一個方向都能良好的分類負荷值，而 RF 模型能有效區分設備的健康狀態且 Y 軸對健康狀態變化的敏感性較高。此外。最後，我們建立互動式儀表板，以更直觀的方式呈現分析結果。透過這些優化，整體系統將具備故障預警與壽命預測能力，有效提升生產效率與設備安全性，充分展現智慧製造的核心價值。

9 程式碼

Github: https://github.com/pochun123/BDA_IoT-based-Virtual-Health-Determination

10 參考文獻

- [1] DUMOULIN, V., AND VISIN, F. A guide to convolution arithmetic for deep learning. *arXiv preprint arXiv:1603.07285* (2016).
- [2] LEI, Y., YANG, B., JIANG, X., JIA, F., LI, N., AND NANDI, A. K. Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical systems and signal processing* 138 (2020), 106587.
- [3] RANDALL, R. B. *Vibration-based condition monitoring: industrial, automotive and aerospace applications*. John Wiley & Sons, 2021.

11 附錄

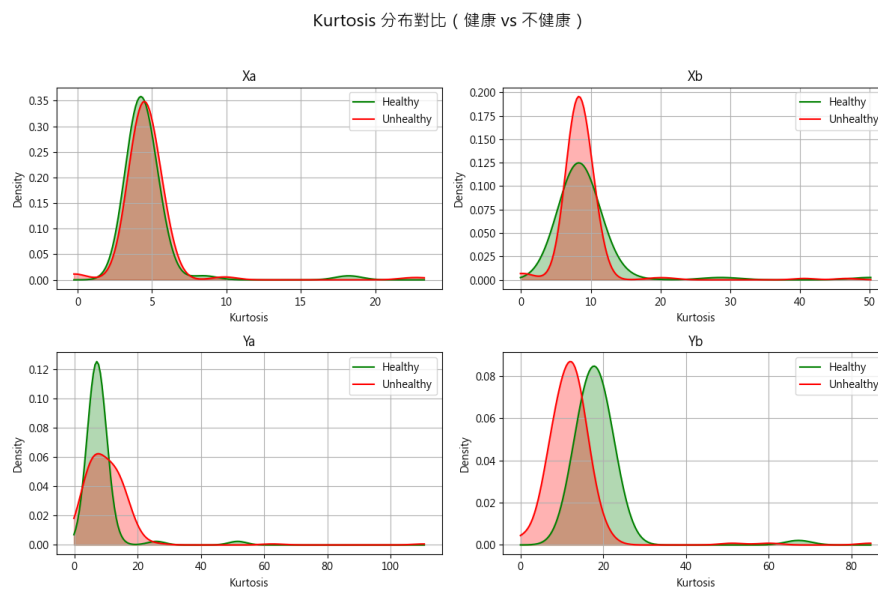


圖 4: Kurtosis 於四方向下健康/不健康樣本分布

RMS 分布對比 (健康 vs 不健康)

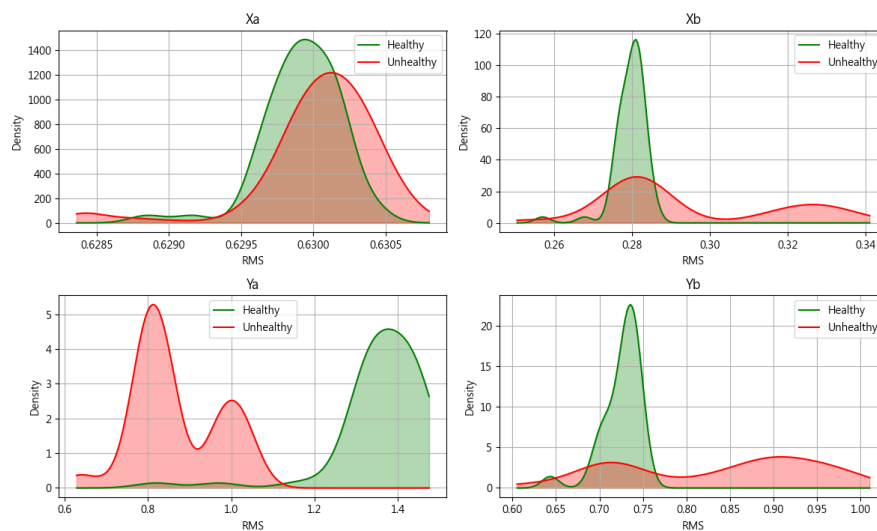


圖 5: RMS 於四方向下健康/不健康樣本分布

RMSF 分布對比 (健康 vs 不健康)

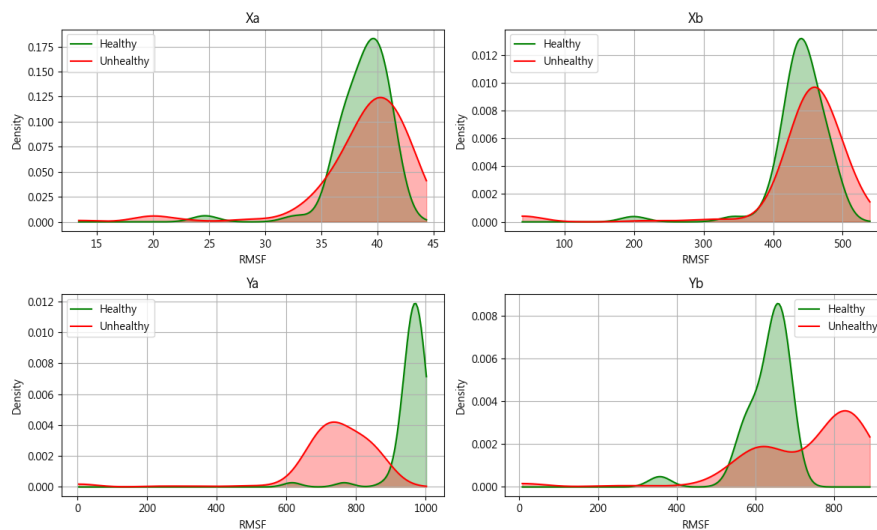


圖 6: RMSF 於四方向下健康/不健康樣本分布

Spectral_Energy 分布對比 (健康 vs 不健康)

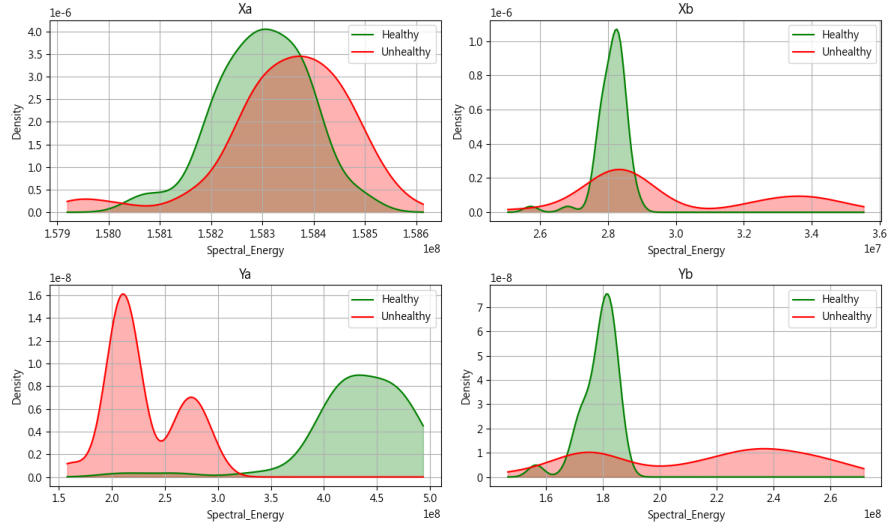


圖 7: Spectral_Energy 於四方向下健康/不健康樣本分布

Spectral_Entropy 分布對比 (健康 vs 不健康)

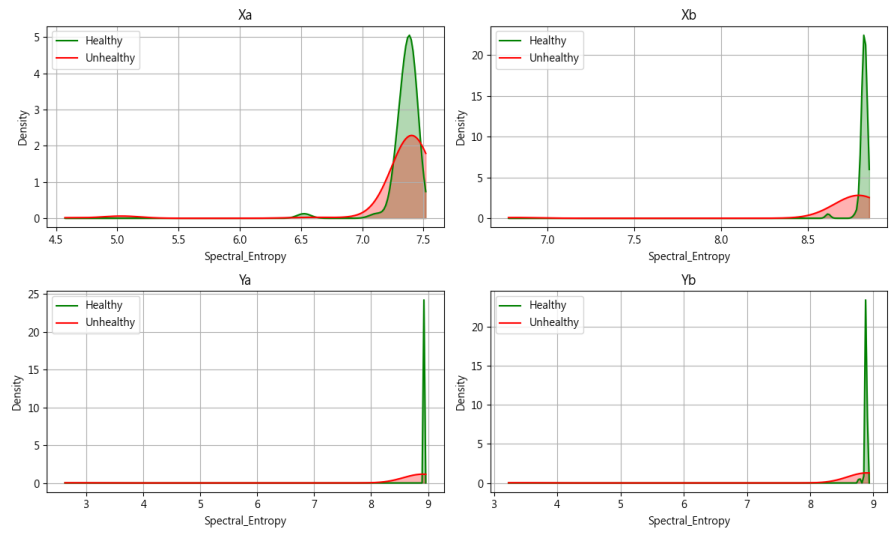


圖 8: Spectral_Entropy 於四方向下健康/不健康樣本分布

Spectral_Kurtosis 分布對比 (健康 vs 不健康)

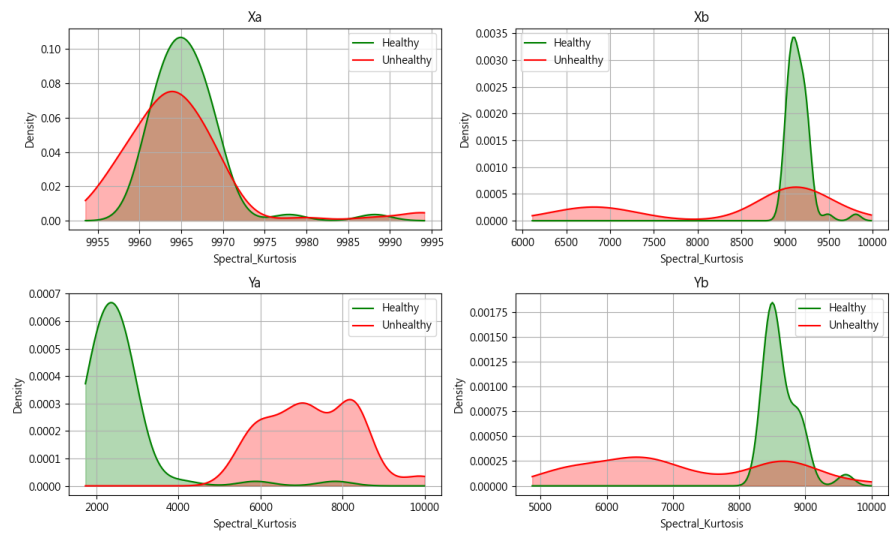


圖 9: Spectral_Kurtosis 於四方向下健康/不健康樣本分布