製造數據科學 Manufacture Data Science

期末專題　書面報告

從聲音訊號到故障預測：利用機器學習實現設備預測性維護

第 G 組

R13725051 陳怡均　R13725057 廖婕妤　B11705056 黃雋亞

B11703063 鄭凱元　B10704026 劉子揚

一、背景與動機

製造設備故障將造成工廠產能下降，而傳統反應性維護之診斷往往依賴人工經驗，可能導致診斷結果不一致或錯誤，降低維修效率與準確性，增加設備故障風險或過度維修。且傳統監控手段不見得能夠辨識所有潛在異常，如設備發出聲音的細微變化，常常是機器已經出現問題後才發現異狀，錯過最佳維護時間而引發嚴重的故障。而預防性維護則難以制定合理的檢修週期，可能導致過度維護，增加不必要的成本。傳統的維修方式也缺乏系統性得收集資料，也無法建立訓練的基準或異常比對，缺乏長期數據也會限制趨勢分析與設備的健康診斷。

因此，運用資料分析技術達到的預測性維護成爲工業 4.0 時代的重要目標。實務上可透過感測器監測設備運行數據，並結合資料分析技術來預測設備潛在故障，進而在故障發生前進行維修，有效減少停機時間並最佳化資源配置。如此一來亦能系統性地收集資料，建立模型訓練基準，以追蹤長期趨勢。

本專案運用深度學習 CNN 與傳統機器學習 Random Forest 等方法建立機械設備所發出音訊之故障偵測模型，並建置能夠接收製造現場資料以實時顯示偵測結果的網頁應用。

二、資料集與研究方法

本專案使用 MAFAULDA - Machinery Fault Database 資料集，資料集內包含透過機械故障模擬器取得之 1951 筆多變量時間序列資料，共有六種故障類型，部分類型包含數種故障元件。每筆資料皆以 50kHz 取樣五秒，共 250,000 列，並各有 8 個特徵欄位，分別是轉速表訊號、下旋軸承的軸向、徑向、切向的加速計、上旋軸承的軸向、徑向、切向的加速計以及麥克風訊號。

本專案將故障類型的標籤進一步擴展爲 10 種，包含正常、水平與垂直的未對準、不平衡與上、下懸軸的保持架故障、外圈故障和滾珠故障，以辨別設備不同元件之異常。資料集內各分類之分佈相對平衡，僅正常狀態之數據較少，各分類資料之標籤及資料筆數如表 1。

表 1：各分類資料之標籤及資料筆數

標籤	資料筆數
正常	49
水平未對準	197
垂直未對準	301
不平衡	333
下懸軸承	
下懸軸承保持架故障	188
下懸軸承外圈故障	184
下懸軸承滾珠故障	186
上懸軸承	
上懸軸承保持架故障	188
上懸軸承外圈故障	188
上懸軸承滾珠故障	137
合計	**1951**

對於 MAFAULDA 資料集，本專案提出了四種模型訓練方法，包括兩個以深度學習 CNN 模型爲基礎的方法與兩個以傳統機器學習模型 Random Forest 爲基礎的方法，兩兩之間之模型訓練流程亦有所差異。

（一） CNN 方法

在 CNN 方法中，聲音資料處理分爲五個主要步驟，目的是將原始訊號轉換爲適合 CNN 模型訓練的特徵圖，並強化模型在實際應用中的表現穩定性與泛化能力。

1. **音訊切片（Segmentation）**：每段長音訊切割爲固定長度的片段（1 秒 = 44100 點），作爲模型的基本輸入單位。這樣可以統一資料格式，也有助於捕捉聲音中短期的異常模式。

2. **資料增強（Data Augmentation）：** 爲了提升模型的穩健性與泛化能力，增加模型對變異的容忍度，本專案在訓練資料中隨機加入三種變形：
    a. **白噪音（add_noise）**：在每段聲音訊號上隨機加入高斯分布的微小擾動，模擬現實中的背景干擾情境，並以參數控制噪音強度，避免過度失眞，提升模型的抗雜訊能力與泛化性。
    b. **音量變化（random_gain）**：模擬不同錄音設備或距離所造成的音量變化，加入隨機音量調整，讓模型能在不同聲音強度下仍準確辨識，提升對現實環境的適應能力。
    c. **時間伸縮（time_warp）**：加入時間伸縮模擬聲音播放速度的自然變化，讓模型能在機器節奏不同的情況下仍能正確辨識。
3. **特徵提取（Feature Extraction）**：將每段 1D 聲音訊號轉換爲 2D 特徵圖（log-Mel spectrogram），輸出 shape 爲 (128, 256, 1)，方便 CNN 進行處理，並保留時間與頻率資訊，有助於模型學習局部結構。
4. **動態特徵（Delta Feature）**：加入一階與二階差分（速度與加速度），幫助模型捕捉聲音變化的趨勢與流動性，識別短暫但關鍵的異常聲音。
5. **特徵標準化（Z-score Normalization）**：對每個樣本進行 Z-score 標準化，消除不同樣本之間的能量差異，讓模型能專注於聲音的相對特徵，而非絕對音量大小。

根據上述資料前處理的結果，本專案分別以下列兩種模型架構來進行訓練：

1. **CNN**
    - **Input 層：**輸入形狀爲 (128, 256, 1) ，爲單聲道的聲音影像輸入。
    - **CNN 層：**建立三層卷積模組，分別使用 32、64 和 128 個濾波器，每層會進行 多個 3x3 濾波器的卷積操作，卷積後會搭配 BatchNormalization 來穩定訓練並加速收斂，再接上 ReLU 非線性激勵函數幫助模型學習非線性關係，最終使用 MaxPooling2D 將空間維度縮小一半，逐步提取局部區域的空間特徵。
    - **Reshape：**將卷積後的特徵圖轉換成 2D 的序列格式，以便後續的 Attention 加權操作。

- **Attention 機制：**模型會從 Reshape 後的序列特徵中，針對每個時間步計算重要性分數，這些分數經過 Softmax 正規化後成爲權重，接著將每個時間步的特徵向量乘上對應的權重，最後使用 reduce_sum 整合成一個加權平均的特徵向量，讓模型能專注在關鍵的序列特徵上。
- **Dense 層：**將 Attention 機制整合後的向量轉換成 256 維，搭配 GELU 激勵函數進行非線性轉換，強化特徵表現力。
- **Dropout 層：**爲了避免過度擬合，使用 Dropout（0.5）隨機丟棄 50% 神經元輸出，提升模型的泛化能力。
- **輸出層：**最後透過 Dense 層輸出與類別數量相同的節點，搭配 softmax 激勵函數，輸出每個類別對應的預測機率。

模型架構如圖 1。

圖 1：CNN 模型架構

2. CNN + BiLSTM

爲了讓模型不僅能分析單一時間點的聲音特徵，也能捕捉聲音訊號在時間序列上的前後關聯性，以進一步提升對聲音類別的判斷準確性。因此本專案在前述 CNN 的架構上加入一層 Bidirectional LSTM（BiLSTM）層，進而從 CNN 擷取的局部圖像特徵中，學習特徵在時間序列上的變化模式，讓模型具備理解聲音在時間上連續性與變化趨勢的能力。

具體來說，本專案在 CNN 卷積層之後，先將卷積後的特徵圖 Reshape 成 2D 的序列格式，接著再串接一層 BiLSTM。此 BiLSTM 層的輸出是 forward 和 backward 各 128 維，共 256 維的特徵向量。同時本專案在模型中設定 return_sequences=True，以保留每個時間步的輸出序列，讓後續能使用 Attention 機制進行加權處理。模型架構如圖 2。

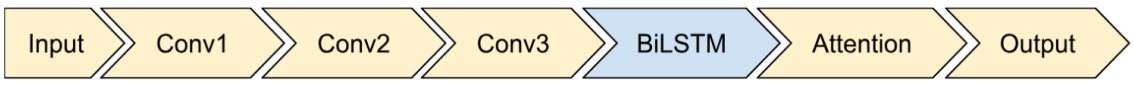


圖 2：CNN + BiLSTM 模型架構

一開始本專案採用資料集中的全部 8 個欄位作爲輸入特徵，包含轉速訊號、加速度感測器（三軸 ×2 組）與麥克風資料，目的是希望整合多種感測器資訊，讓模型能同時學習機器的運轉狀態與異常行爲，進而提升分類的全面性與泛化能力。

然而，實驗結果顯示，在使用所有欄位進行訓練的情況下，模型的分類準確率僅約 60%，進一步分析後發現，可能原因在於不同感測器的數據特性差異較大（如振動訊號與聲音訊號），使得模型難以同時擷取穩定且一致的特徵；再加上多通道資料難以整齊對齊，導致特徵圖結構混亂，進而影響模型的判斷能力與學習焦點。

因此，本專案調整策略，聚焦使用第八欄（即麥克風聲音訊號）作爲模型輸入，並且搭配 log-Mel 時頻圖轉換與 CNN 架構進行訓練，讓模型能專注學習聲音中的關鍵模式。去除其他感測器帶來的干擾後，模型表現明顯提升，最終實驗結果也顯示分類準確率有顯著改善。

（二）Random Forest 方法

Random Forest 爲傳統機器學習方法之模型，該模型能夠捕捉非線性之特徵，並且由於是 tree-based ，對於特徵範圍不敏感，不需要再進行標準化。此外，使用 Random Forest 訓練模型亦提供了良好的解釋性，能夠識別重要特徵作爲篩選與重新訓練之依據。本專案提出了兩個透過 Random Forest 的訓練方法。

1. Naive Random Forest：以各欄位統計量作爲特徵

將每一筆資料（每筆資料爲一個含有 8 個原始特徵的 .csv 檔）的八個感測器訊號欄位，每個欄位皆提取平均值、標準差、最大值、最小值及中位數等五個統計量，最終以 40 個特徵作爲模型輸入。

進行特徵工程後，將資料以 8 比 2 切分爲訓練集及測試集，分割後再針對 40 個特徵進行標準化，隨後用以訓練模型並測試。

2. 採用統計與頻率特徵，結合資料平衡與特徵篩選

統計特徵部分，此方法提取 8 個感測器訊號之平均值、標準差、最大值、最小值、均方根、偏度、峰度及峰值因數等 8 個統計量，共產生 64 個特徵。此方法亦提取短時距傅立葉轉換之頻率特徵，參考 Matheus A. Marins 等 (2017) 發表之期刊文章，不同的諧波可能可以偵測不同的故障類型，故取轉速表以外的七個感測器訊號於

旋轉頻率 1 倍、 2 倍及 3 倍之頻譜幅值平均 [1]，共產生 21 個特徵。加上由轉速表計算之旋轉頻率，共有 86 個特徵。

此方法將資料集以 8 比 2 切分爲訓練集及測試集，另也以 5 個 fold 進行交叉驗證。針對訓練集以及用來訓練之 fold，此方法採用過採樣方法 SMOTE 進行資料平衡，以避免模型未能完整學習資料較少之類別（如 normal 僅有 49 筆資料），資料平衡前後之 t-SNE 降維資料分布如圖 3。

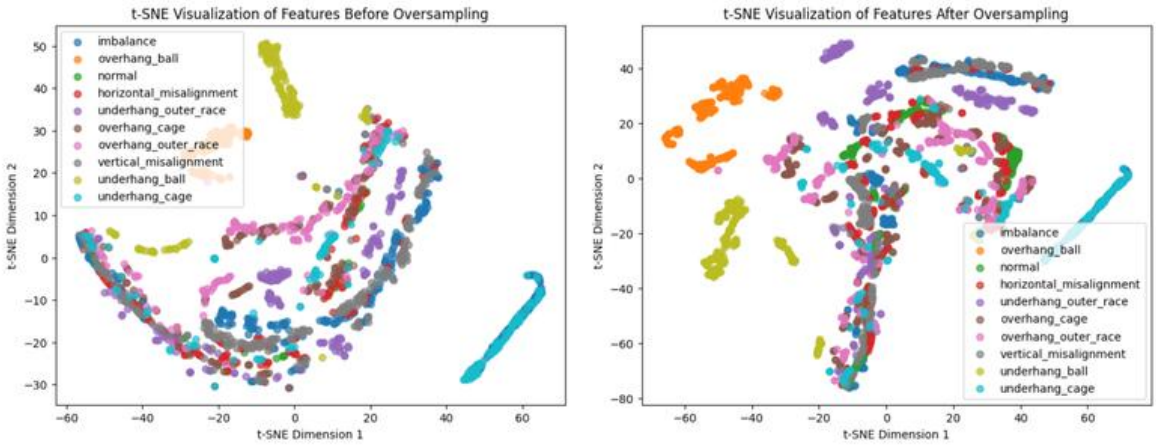


圖 3：資料平衡前後之 t-SNE 降維資料分布

由於一開始提取的特徵數量較多，此方法亦篩選重要特徵，避免特徵數量過多導致的維度詛咒問題。首先以所有特徵訓練模型，並篩選出 Permutation Importance 爲正值的 22 個特徵訓練最終模型。各特徵之 Permutation Importance 如圖 4 所示，可觀察出多數特徵對模型貢獻度較低。

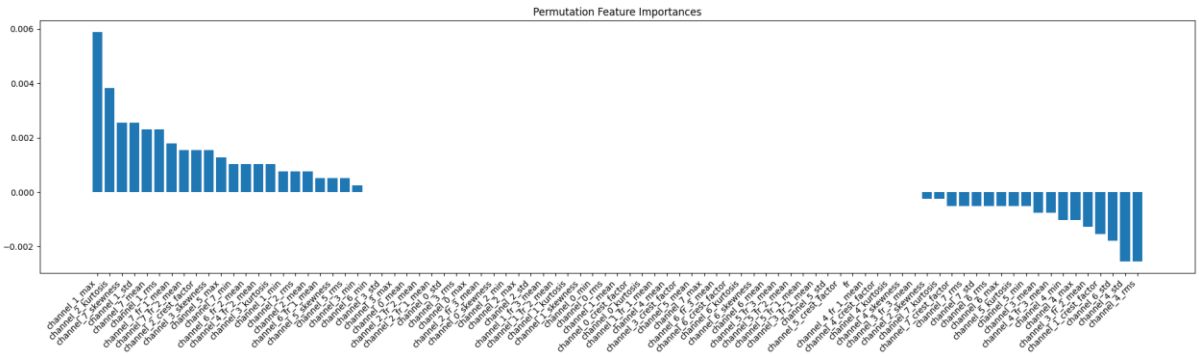


圖 4：各特徵之 Permutation Importance

三、結果分析

在一開始的 CNN 結果中，無論有沒有加入 BiLSTM，在多通道輸入的設定下，模型同時接收來自不同感測器的訊號（轉速訊號、加速度計三軸 × 兩組以及麥克風聲音訊號），理論上能整合多源資訊以提升分類效能。然而，實驗結果顯示：準確率皆只有來到 60% 左右（圖 5），根據混淆矩陣（圖 6）可以發現類別如 0、1、2、5、9 難以被正確分類，且學習表現不穩，泛化能力有限，代表模型在處理多通道輸入時，可能受到資料間分佈不一致、尺度不同、雜訊干擾等問題影響，使得特徵難以有效融合，導致學習效率不佳。

Classification Report:

	precision	recall	f1-score	support
0	0.4681	0.3667	0.4112	60
1	0.7727	0.5667	0.6538	60
2	0.6444	0.4833	0.5524	60
3	0.5222	0.7121	0.6026	66
4	0.9474	1.0000	0.9730	54
5	0.6538	0.5667	0.6071	60
6	0.7937	0.8333	0.8130	60
7	0.9206	0.9667	0.9431	60
8	0.6049	0.8167	0.6950	60
9	0.5172	0.5000	0.5085	60
accuracy			0.6783	600
macro avg	0.6845	0.6812	0.6760	600
weighted avg	0.6803	0.6783	0.6723	600

Classification Report:

	precision	recall	f1-score	support
0	0.4839	0.2500	0.3297	60
1	0.5625	0.7500	0.6429	60
2	0.4630	0.4167	0.4386	60
3	0.5122	0.6364	0.5676	66
4	0.9444	0.9444	0.9444	54
5	0.5111	0.3833	0.4381	60
6	0.7222	0.8667	0.7879	60
7	0.9355	0.9667	0.9508	60
8	0.6250	0.6667	0.6452	60
9	0.5179	0.4833	0.5000	60
accuracy			0.6333	600
macro avg	0.6278	0.6364	0.6245	600
weighted avg	0.6234	0.6333	0.6207	600

圖 5：採用多通道 CNN 及 CNN + BiLSTM 模型之分類結果報告

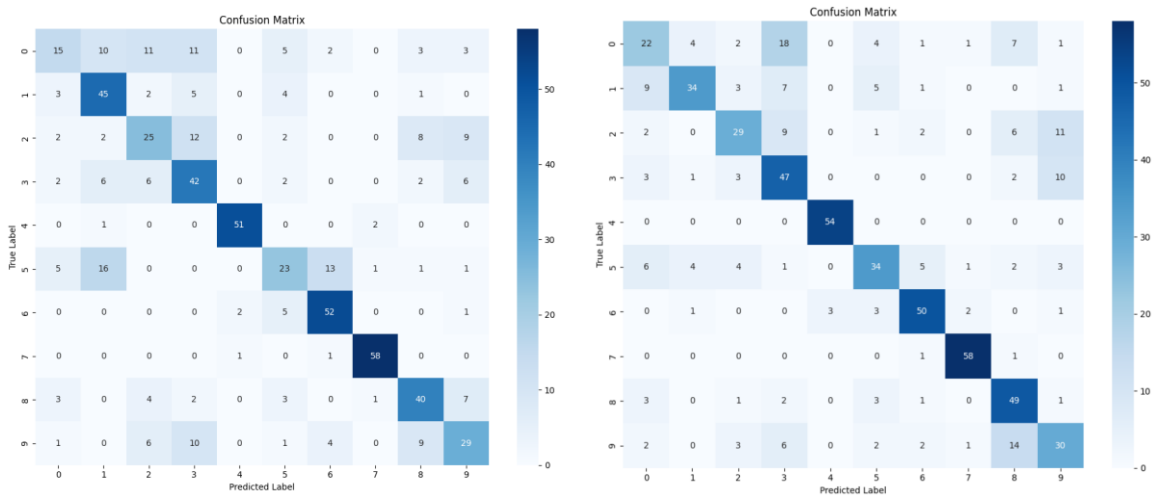


圖 6：採用多通道 CNN 及 CNN + BiLSTM 模型的混淆矩陣

在後續實驗中，改採用「僅第八欄的麥克風聲音訊號」作爲輸入，經 log-Mel 頻譜圖轉換後輸入 CNN 模型，目標是去除其他感測器的干擾，聚焦聲音模式學習。實驗結果顯示，採用單通道輸入能顯著提升模型分類表現，CNN 模型的準確率提升至 77%（圖 7），進一步再加入一層 BiLSTM 捕捉聲音序列的時間關聯性，準確率更能提升到 85% 左右。透過混淆矩陣 （圖 8）可以發現除了第 0、1 類別仍較難被準確分類外，其餘類別幾乎都能有效區分，表示聲音訊號具備足夠的判別度，且排除其他感測器干擾能讓模型更穩定地學習。

Classification Report:

	precision	recall	f1-score	support
0	0.5714	0.4667	0.5138	60
1	0.6111	0.5500	0.5789	60
2	0.7067	0.8833	0.7852	60
3	0.8971	0.9242	0.9104	66
4	0.9130	0.7778	0.8400	54
5	0.8571	0.9000	0.8780	60
6	0.7879	0.8667	0.8254	60
7	0.7321	0.6833	0.7069	60
8	1.0000	0.8833	0.9381	60
9	0.6714	0.7833	0.7231	60
accuracy			0.7733	600
macro avg	0.7748	0.7719	0.7700	600
weighted avg	0.7746	0.7733	0.7707	600

Classification Report:

	precision	recall	f1-score	support
0	0.5833	0.4667	0.5185	60
1	0.5469	0.5833	0.5645	60
2	0.7536	0.8667	0.8062	60
3	0.9538	0.9394	0.9466	66
4	1.0000	0.8889	0.9412	54
5	0.9231	1.0000	0.9600	60
6	0.9355	0.9667	0.9508	60
7	0.9062	0.9667	0.9355	60
8	1.0000	1.0000	1.0000	60
9	0.9818	0.9000	0.9391	60
accuracy			0.8583	600
macro avg	0.8584	0.8578	0.8562	600
weighted avg	0.8580	0.8583	0.8563	600

圖 7：採用單通道（聚焦第八欄）訓練後的 CNN 及 CNN + BiLSTM 模型分類結果報告

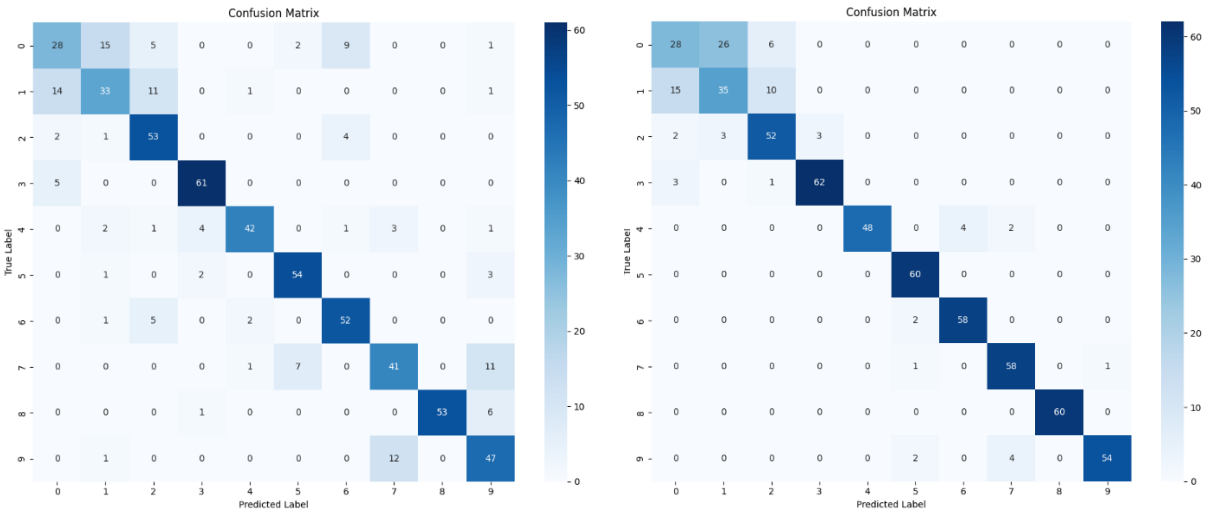


圖 8：採用單通道（聚焦第八欄）訓練後的 CNN 及 CNN + BiLSTM 模型混淆矩陣

至於 Random Forest 方法，以 Naive Random Forest 模型測試結果簡單平均 Accuracy 高達 0.98（圖 9），進一步利用混淆矩陣可以確認模型預測並沒有盲猜的情形（圖 10）。

	precision	recall	f1-score	support
normal	0.90	0.90	0.90	10
imbalance	0.97	0.97	0.97	67
horizontal_misalignment	0.95	0.95	0.95	39
vertical_misalignment	1.00	0.98	0.99	60
overhang_ball	1.00	1.00	1.00	27
overhang_cage	0.95	1.00	0.97	38
overhang_race	1.00	0.97	0.99	38
underhang_ball	1.00	1.00	1.00	37
underhang_cage	1.00	1.00	1.00	38
underhang_race	1.00	1.00	1.00	37
accuracy			0.98	391
macro avg	0.98	0.98	0.98	391
weighted avg	0.98	0.98	0.98	391

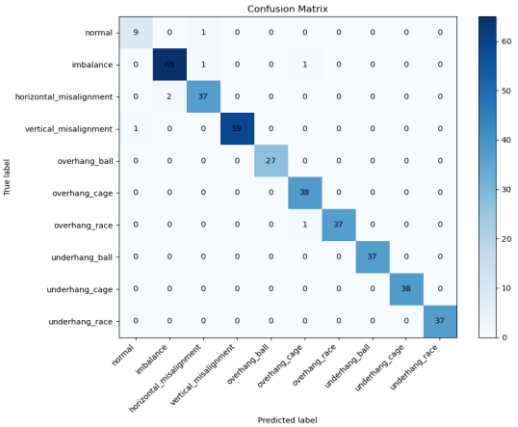
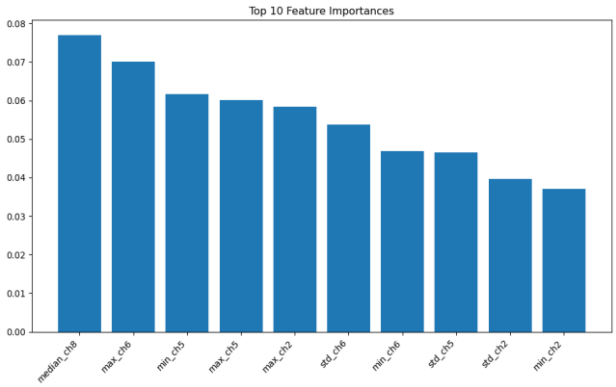


圖 9：Naive Random Forest 之分類報告　圖 10：Naive Random Forest 之混淆矩陣

從特徵重要性分析可以發現最重要的特徵爲 median_ch8，即麥克風訊號的中位數（圖 11），顯示麥克風音訊特徵確實有助於機器故障的判斷。圖 12 顯示在不同的測試集比重下模型的準確性，可以發現在達到一定的訓練資料量後模型表現穩健。



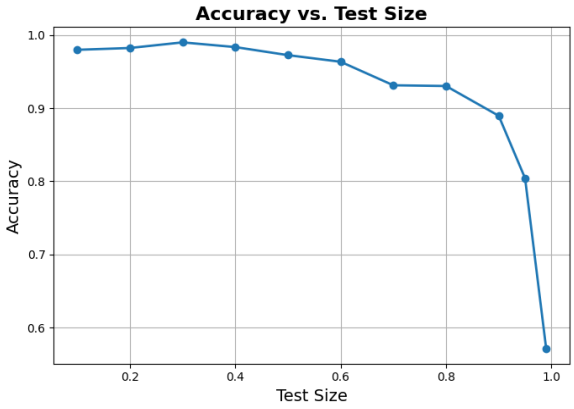


圖 11：重要性前十名之特徵　圖 12：不同測試集比重下之 Accuracy

在 Random Forest 採用統計與頻率特徵並結合資料平衡與特徵篩選之方法中，以最終模型預測測試集之 Accuracy、Precision、Recall 及 F1-score 均達 99%，混淆矩陣如圖 13 所示。以 5 個 fold 交叉驗證之平均 Accuracy 爲 91.80%、Precision 爲 92.55%、Recall 爲 91.18%、F1-score 爲 91.08%，成效良好。

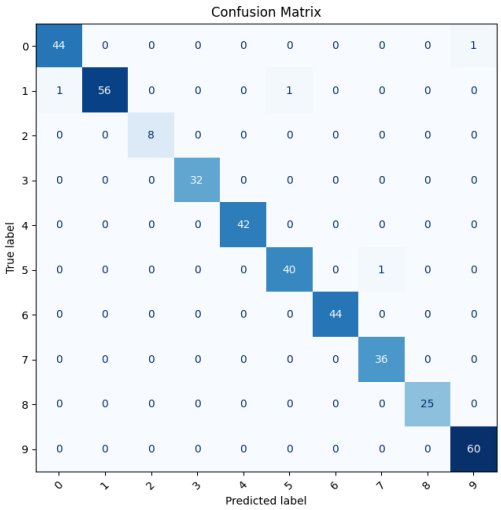


圖 13：Random Forest 採用統計與頻率特徵訓練最終模型之混淆矩陣

爲因應製造現場之使用需求，本專案亦建立能夠接收資料並即時預測的網頁應用，該應用會監聽 sensor_data 資料夾，並配合感測器資料五秒建立一個檔案的週期，每五秒抓取資料夾中的檔案，在有新的 csv 檔案加入時即時偵測設備異常種類。預測結果會被儲存至資料庫中，並顯示過去之偵測結果、預測爲該類別的機率等資訊在網頁介面上，最近一個檔案的預測結果則會顯示在畫面上方，並以背景顏色標示正常與否。使用者亦可設定只顯示異常之結果，避免佔多數的正常結果占用版面。網頁應用之介面如圖 14。

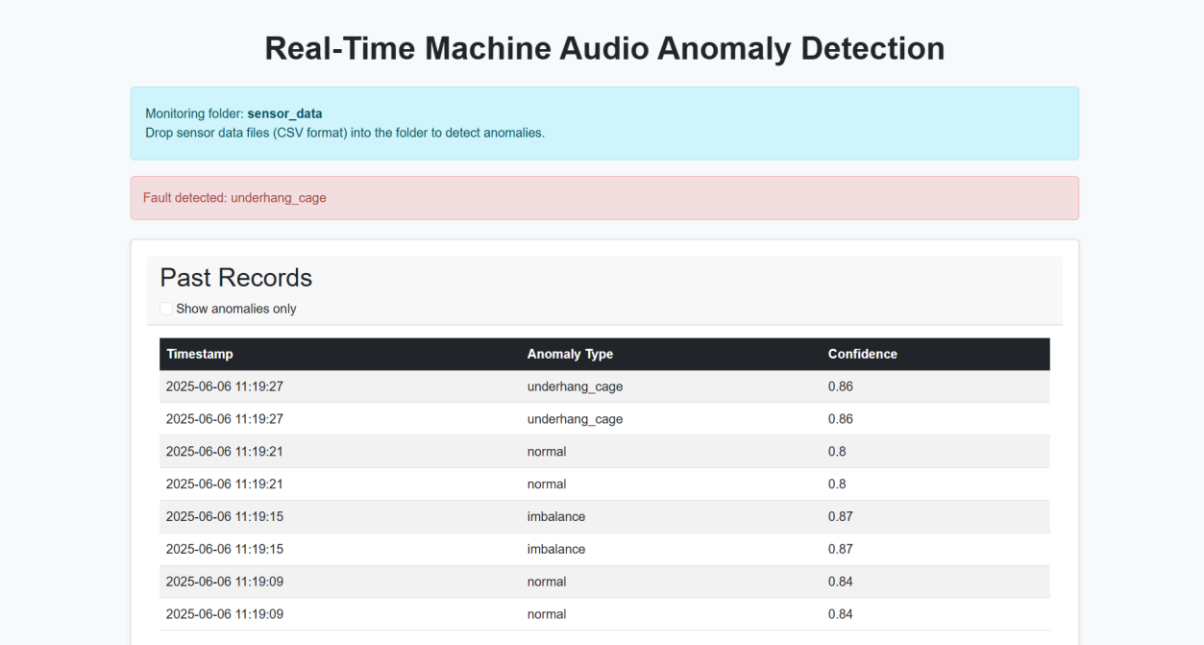


圖 14：網頁應用介面

四、討論與結論

本專案提出了四種模型訓練方法，包括兩個以深度學習 CNN 模型爲基礎的方法與兩個以傳統機器學習模型 Random Forest 爲基礎的方法。其中深度學習之 CNN 模型訓練時間較長、運算資源需求也較高，然其預測成效卻不及傳統機器學習之 Random Forest 模型，故可推論在目前之資料量及任務需求下，可透過較低的時間與資源需求獲得更好訓練成效的 Random Forest 模型是較佳的選擇。

實驗結果顯示，Random Forest 模型在多種設備異常情境下展現出優異的表現，準確率高達 98% 以上，顯示其具有穩定且可靠的故障辨識能力。本專案建置之網頁應用能夠搭配模型，於異常發生前即時發出預警，通知工程人員及早介入處理，有效降低產線停擺的風險。此外，該模型可取代傳統定期巡檢所需的人力，降低人力成本，使工程人員得以將資源集中於高風險設備的維護工作上。更重要的是，透過預測性維護策略，企業得以主動安排設備維護時程，及早更換零件或調整參數，以降低非預期停機與產線中斷所造成的損失，同時延長設備壽命並提升整體運作穩定性。

目前所使用的資料集係由機械故障模擬器所產生，雖已具備良好成效，惟爲提升模型於眞實應用環境中的效能與適用性，未來將規劃收集實際感測器資料以進行模型驗證與最佳化。此外，模型目前僅聚焦於三種關鍵零組件之故障偵測，未來可進一步擴展至更多類型的關鍵元件，提升故障診斷的完整性與精準性。同時，亦可導入自動化的模型效能監控與再訓練機制，確保模型能持續適應新興資料特徵，進而維持長期的準確度與實用性。

五、參考文獻

[1] Matheus A. Marins, Felipe M.L. Ribeiro, Sergio L. Netto, Eduardo A.B. da Silva (2017), Improved similarity-based modeling for the classification of rotating-machine failures, Journal of the Franklin Institute, Volume 355, Issue 4, 2018, Pages 1913-1930.