

# 2023 年臺灣國際科學展覽會 優勝作品專輯

作品編號 190002

參展科別 電腦科學與資訊工程

作品名稱 以深度學習進行心音及高血壓關聯性之研究

得獎獎項 二等獎

土耳其音樂科學工程博覽會(Buca IMSEF)代表  
義大利科學博覽會(FAST)代表

就讀學校 國立新竹科學園區實驗高級中等學校

指導教師 劉奕汶

作者姓名 陳亭儒、王允芊

關鍵詞 高血壓、心音時序頻譜圖、圖像分類模型

## 作者簡介



大家好，我是實驗高中的陳亭儒。我從國二科展時開始接觸 AI，並發展出濃厚的興趣。高中時認識了允芊，一起決定將國二時的科展主題延伸到醫療領域。因緣際會下我們參加了未來之星的培訓，並從中獲得了許多啟發。我們在研究專題的路上跌跌撞撞地前行，儘管遇到了無數阻礙，卻也讓我見識到 AI 的無限潛能並期待著 AI 在各領域的革命性發展。

大家好，我是王允芊，在高一時認識了亭儒，也開始接觸 AI，因此決定一起做科展。過程中，我們很幸運的參加了未來之星的培訓，一起在這一路的磕磕絆絆中成長與學習，也讓我發現越來越多 AI 的有趣之處。雖然做專題一直讓我們陷入無助又焦慮的 emo 循環，但在實驗中的有趣發現仍能促使著我們繼續堅持向前邁進。

## 摘要

2019 年衛生福利部死因統計資料顯示和高血壓有高度相關的心臟疾病、腦血管疾病和高血壓性疾病皆在十大死因之列[15]。本研究提出以深度學習對心跳聲的時序頻譜圖進行訓練與分析的研究方法，應用此方法我們能以 Convolution Neural Network(CNN)模型從受測者心跳聲預測出其血壓層級。CNN 一般用於圖像分類，但在此研究中我們以此來分析心跳聲。本研究發現利用僅萃取第二心音的資料庫訓練效果較佳，並透過熱圖分析注意到模型對特定頻率域較為重視，在後續實驗中更進一步發現 0~200 Hz 和 400~600 Hz 在判斷高血壓時扮演重要角色。同時，我們也成功應用此方法，區分出長期高血壓和運動高血壓，證明心血管的結構改變在時序頻譜圖上有對應特徵。若應用於穿戴型裝置持續監控心跳聲，就能隨時追蹤使用者的血壓層級的變化，有異常便能盡早就醫，避免憾事發生。

## Abstract

In 2019, The cause of death statistics published by the Health Promotion Administration (HPA) indicated that heart disease, cerebrovascular disease, and hypertensive diseases, which are highly related to hypertension, are all among the top 10 causes of death [15]. In this report, we propose a method of using spectrograms of heartbeats as a dataset, and by utilizing deep learning models, blood pressure levels can be predicted. We use Convolution Neural Network, which is typically used in image classification, to analyze heart sounds. We discovered that using a dataset consisting of the second heart sound had a better result, and noticed that the model places more emphasis on specific frequency domains by using heatmap. In the subsequent experiments, it was further found that 0-200 Hz and 400-600 Hz play an important role in detecting hypertension. Moreover, we distinguished exercise-induced hypertension and long-term hypertension, which proves that structural differences of hearts and vessels have corresponding features on spectrograms. This method is suitable for wearable devices monitoring user heartbeats and tracking blood pressure. Early alarm for medical treatment would then become easy and achievable.

## 壹、研究動機

在生活周遭，我們經常聽到長輩有高血壓，需要定期吃藥並追蹤血壓值。根據世界衛生組織的統計，全世界約有 12.8 億人有高血壓，其中有約 7.2 億人沒有得到合適的治療[8]。進一步調查發現，高血壓可能導致心血管疾病、腦中風、糖尿病、腎臟病等重大慢性病，是不容忽視的危險因子[9]。

現在大多使用血壓計來量測並追蹤血壓，然而血壓計無法連續地測量，需要仰賴人們自我提醒來追蹤，且加壓部位若有傷口可能造成不適。而現有的穿戴式量測裝置多藉由光學測量血管壁中血流量變化，因此外界環境、膚色等等都很容易影響到測量結果，誤差甚至可達 35% [13]。故在本研究中，我們希望運用深度學習技術，透過心跳聲來判斷血壓的層級，將來可運用在穿戴式裝置上，讓人們可以隨時監測自己的血壓狀況。除此之外，本研究是以心跳聲轉為頻譜圖後使用圖像分類模型進行訓練，因此也可以透過可視化模型提取特徵來了解哪部分的心跳聲頻率和血壓具有關聯性，亦可對相關學術研究貢獻一己之力。

## 貳、研究目的

- 一、驗證以心音時序頻譜圖結合圖像分類模型的可行性。
- 二、建立心音頻譜圖資料庫，用 CNN 模型判斷受測者的血壓層級。
- 三、比對心音成分對模型訓練成果的影響。
- 四、比對模型層數對模型訓練成果的影響。
- 五、比較心音中不同頻率段對模型訓練成果的影響
- 六、比對受測者心音差異對模型訓練成果的影響。
- 七、建立模型辨別長期高血壓與運動後高血壓心音之差異。

## 參、文獻回顧

### 一、短時傅立葉變換技術探討

短時傅立葉變換(Short-time Fourier Transform, STFT)是一種建立在傅立葉變換上的衍生技術，它將長時間的聲音信號分為較短且等長的數個信號再分別計算短時間的傅立葉變換，是時頻分析中的重要工具。

此轉換可將一段聲音信號轉換為能夠紀錄隨時間變化的頻率與聲音大小資訊的時序頻譜圖(Spectrogram)，圖中會以縱軸表示頻率成分、橫軸表示時間、明暗表示聲音強度。

在使用短時傅立葉變換時，可自行調動窗的長度來決定時間分辨率與頻率分辨率。若窗長越長，擷取的聲音信號也會越長，經過變換的頻率分辨率越高，而時間分辨率越差，反之亦然。故時間分辨率與頻率分辨率無法兼得，需依照自己的需求進行取捨，調出最適合的窗長。

短時傅立葉轉換的計算方式，會先將一個函數乘以窗函數，再進行一維的傅立葉變換，並且窗函數會隨著時間挪移，得到一系列的傅立葉變換結果，並將此結果排開成為二維表象，數學定義可表示為：

$$X(t, f) = \int_{-\infty}^{\infty} w(t - \tau)x(\tau)e^{-j2\pi f\tau} d\tau.$$

上式中， $w(t)$  為窗函數， $x(t)$  為待分析的訊號。

### 二、CNN 技術探討

卷積神經網絡(Convolution Neural Network, CNN)是深度學習領域中的重要演算法，其結構類似於人腦神經元的連接模式，靈感來自視覺皮層的組織，在影像辨識方面應用極為廣泛。透過卷積層、池化層和全連接層，CNN 能夠利用共享權重降低參數量，並同時保留圖

片特徵。而模型的卷積層、池化層和全連接層數量會造成不同的訓練結果，因此後續實驗也會針對此部分進行調整。

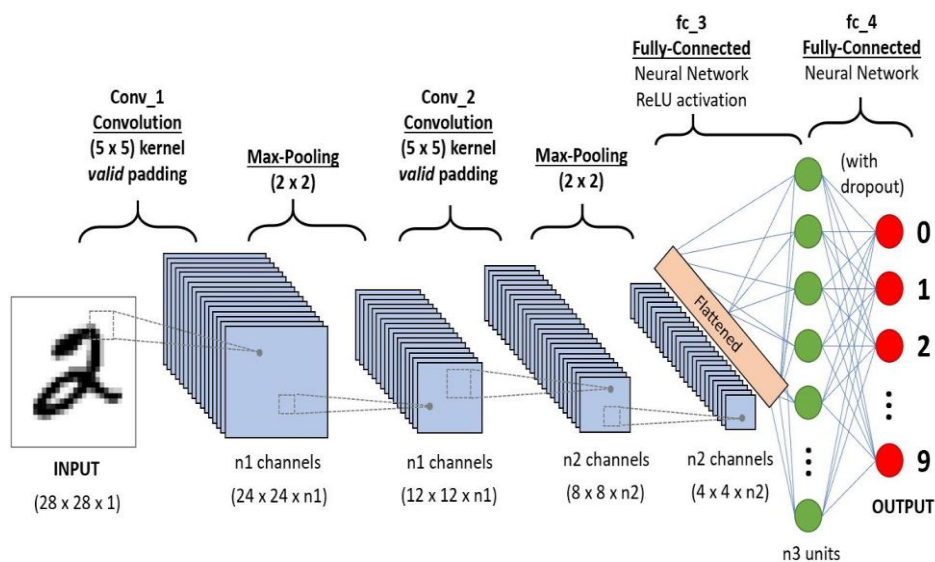


圖 1 CNN 模型的運作方式[7]

### 三、高血壓

根據世界衛生組織(World Health Organization, WHO)，高血壓是一種嚴重的醫療狀況，會顯著增加心臟、大腦、腎臟及其他疾病的風險[8][16]。高血壓的症狀有頭痛、心律不整，嚴重者嘔吐、胸痛甚至中風、猝死[9]。根據 WHO 的標準，在不同天量測，若兩天收縮壓皆大於等於 140mmHg 或舒張壓大於等於 90mmHg 則判定為高血壓。

高血壓會導致血液流速較快，撞擊血管壁的力道也會較大，而導致心跳聲的響度較大。另一方面，血壓高也會導致動脈的復位力較大，使震動的頻率較大。每一次的心音週期包含四個心音，一般能聽到的是第一及第二心音。第一心音(S1)是心臟收縮的開始，主要由心臟內二尖瓣和三尖瓣突然關閉而血液突然衝擊血管產生渦流所形成，其時間較長且頻率較低，心臟收縮力越強則第一心音越響；第二心音(S2)則是心臟舒張期的開始，主要由動脈瓣突然關閉，血液突然減速而造成動脈根部一起震動所形成，其時間較短且頻率較高。而如果血壓

高，動脈壁需要給予的復位力會增加，進而導致其振動頻率增加[14]。因此與第一心音相比，第二心音與高血壓存在更高的相關性[11]。

高血壓有多種成因，可分為原發性高血壓和續發性高血壓兩種。原發性高血壓即是不明原因造成的高血壓，可能是遺傳、飲食、壓力、缺乏運動等原因。續發性高血壓則是有特定高血壓的誘因，如腎臟疾病、先天動脈血管疾病、內分泌失調，或特定藥物的副作用等等[12]。上述之高血壓屬於長期血壓過高的狀況，然而運動也能短期性造成高血壓。在運動時，肌肉對於氧氣的需求量會提升，因此交感神經會變得活躍，心跳也會加速以使心臟輸出量增加。由於心跳頻率和每次心跳的輸出率均增加，血壓因此上升[4][5]。

## 肆、研究方法與過程

### 一、研究架構與方法

由於深度學習的相關技術一日千里，本研究嘗試以一概念證明專題(proof of concept)，驗證以心音時序頻譜圖結合圖像分類模型的可行性，並提出如下的研究架構與方法，作為後續研究改良與技術演進的框架。此研究方法分為四個階段：

- (一) 心音採集
- (二) 資料處理
- (三) 模型訓練及評估
- (四) 熱圖分析

本科展具體操作過程如圖 2 所示。首先，我們徵求受試者讓我們量測血壓以及錄製心音，再將錄製到的心音音檔中的單次心跳聲一一擷取，並轉為時序頻譜圖，且利用檔名方式記錄血壓資訊。接著，將處理好的資料集放入 CNN 模型中做訓練，根據模型訓練狀況進行參數調整。訓練完成後再進一步進行心音和高血壓關聯性的分析。

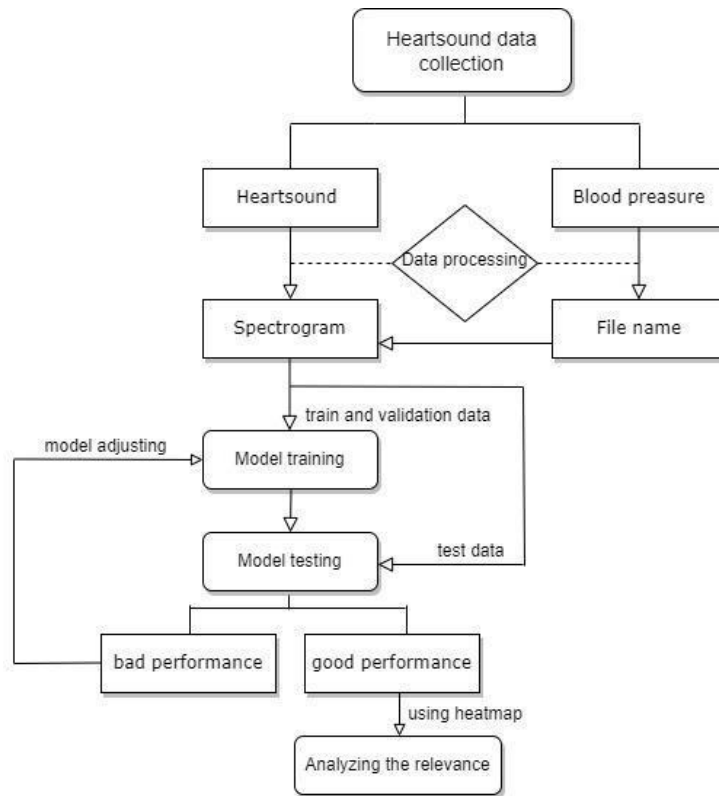


圖 2 實驗流程圖

## 二、研究設備與器材

### (一) 量測設備：

1. 聽診器：Spirit 專業級懸浮式膜面雙面聽診器。蒐集受測者心音。
2. 麥克風：JPB 領夾式 Type-C 麥克風。塞入聽診器軟管中，錄製聽診器心音。
3. 指尖型血氧機：Pulse Oximeter 指夾式血氧檢測機。測量運動後心率做為參考。
4. 血壓計：Rossmax 優盛電子血壓計 CF175f (國、台語語音型)。量測血壓以做資料標記。

### (二) 開發環境：

1. 程式語言：Python
2. 模型訓練與資料處理：



(1)Keras :

Keras 為一種開源神經網路庫，也可以說是開放的高階深度學習程式庫，為建立在 TensorFlow 上的高層 API。

(2)TensorFlow :

TensorFlow 為 Google 所提供的開源深度學習框架，可以支援各種深度學習的演算法。

(3)NumPy :

NumPy 為 Python 的擴充程式庫，主要處理多維度的陣列運算，且通常運用在大量的資料處理上。

(4)OS :

OS 為 Python 內建中與檔案系統相關的模組，可對目錄或檔案進行操作。

3. 圖片生成與處理 :

(1)Matplotlib.Pyplot :

Pyplot 為 Matplotlib 的子庫，是常用的繪圖模組，用來繪製 2D 的圖表。

(2)PIL.Image :

PIL(Pillow)是 Python 中的影像處理套件，而 Image 為 PIL 中最常見的模組，其中包含了幾乎所有對影像進行基礎操作的功能。

(三) 訓練環境：Google Colab

Google Colab 是一個基於 Jupyter Notebook 的線上 Python 的雲端執行環境，它提供免費的 GPU 算力，且可以連接 Google 雲端，獲取雲端上的資料，加上它擁有共同編輯的功能，因此我們選擇使用 Google Colab。

#### (四) 切檔程式：Audacity

此程式為音訊編輯軟體，提供音檔的剪貼、轉檔、正規化、除噪等功能。

### 三、研究流程

#### (一) 心音採集

實驗場景如圖 3 所示，受測者會先在 1 號位置上詳細閱讀受測者知情同意書。若願意參與實驗，則移動至 2 號位置量測血壓，最後至 3 號位置錄製心跳。此實驗在隔音良好的教室進行，以確保資料品質。



圖 3 心音採集場景

本科展依據國內人體研究法、醫療法等相關法規規範，提供「受測者知情同意書」，內詳述研究目的、內容，研究對受測者的益處及潛在風險，皆取得受測對象同意並簽署。若受測者於受測期間希望退出實驗，則立刻停止實驗，並銷毀其個人資訊，且收集到的資料將會以編碼方式做處理，以達到去連結的效果。指導人員亦滿足最近六年研習醫學倫理課程九小時以上之要求。

本科展預計徵求 50 名受測者，每名受測者約錄製 30 秒的心音音檔，其中約包含 20~30 次心音資料，預計總資料數約 1000 筆資料。

本科展由於實驗需求，需要靜坐狀態下量測的血壓及運動後量測血壓兩種資料。受測者首先會在圖 3 的 1 號位置詳細閱讀受測者知情同意書，若願意參與實驗，才會進行後續研究流程。以下針對這兩種資料的蒐集方式分別進行詳細的說明。

靜坐狀態下量測血壓之資料蒐集流程：

- 1.使受測者靜坐於 2 號位置有靠背的舒適椅子 5 分鐘後，量測其血壓，量測期間受測者須保持沉默，且不可觀看機器測量之數值。
- 2.移動至 3 號位置，將研究員改造之可錄製心音的聽診器給予受測者，使其將聽診器的鼓面按壓在胸骨左側邊緣的第二到第三根肋骨之間，並由研究員操作電腦，以錄製心音。

運動後量測血壓之資料蒐集流程：

- 1.使受測者運動後，以指尖型血氧機監測其心律，確認已達運動標準則立即在 2 號位置量測其血壓。根據三軍總醫院所發布的運動心率計算方式，我們將每分鐘心跳達 120 次以上作為運動標準[10]。且量測期間受測者須保持沉默，且不可觀看機器測量之數值。
- 2.移動至 3 號位置，將研究員改造之可錄製心音的聽診器給予受測者，使其將聽診器的鼓面按壓在胸骨左側邊緣的第二到第三根肋骨之間，並由研究員操作電腦，以錄製心音。

## (二) 資料處理

本科展的資料處理分為兩個部分，心音擷取與時序頻譜圖的轉換。

### 1.心音擷取

本科展依據高血壓的成因，將心音音檔分為第一加第二心音(S1+S2)和第二心音(S2)，並使用 Audacity 程式手動擷取心音音檔，切檔流程如下：

(1)將心音音檔匯入應用程式 Audacity，提取心音與心音之間的空白時段的噪音特徵，再利用提取的噪音特徵做音檔的噪音抑制。

(2)將心音音檔中的 S1+S2 擷取後存為.wav 檔，再將其中的 S2 單獨擷取出來存為.wav 檔。

## 2.時序頻譜圖的轉換

本科展中將以短時傅立葉轉換，將擷取後的音檔利用程式批量轉換為時序頻譜圖。

以下為時序頻譜圖的轉換流程：

(1)以迴圈方式利用 fastai\_audio 程式庫中，離散傅立葉轉換(Fast Fourier Transform, FFT)的短時傅立葉變換(Short-time Fourier transform, STFT)將.wav 檔轉成時序頻譜圖。

(2)經轉換後發現圖片會被自動加上白框，為避免白框影響訓練結果，因此利用 plt.axis('off')、bbox\_inches='tight'與 pad\_inches=0 將外圍白框去除。

(3)進一步將時序頻譜圖轉換成 217\*223\*4 的.png 檔儲存，當作本科展深度學習系統的訓練資料集。

## 3.資料命名方式

本科展的錄音檔名格式如下：P1D0S1。

其中，P 為受測者編號，D 為高血壓分級，S 為擷取出的第幾個音檔，依據此方式命名，便可直接從檔名了解該資料的屬性。

其中，血壓分級方式是根據美國心臟學會(American Heart Association, AHA)於 2019 年發布最新的高血壓標準(The New Blood Pressure Guideline)[1]。以 D0 代表血壓正常；D1 代表血壓偏高；D2 代表第一期高血壓；D3 代表第二期高血壓以及高血壓危險期。

表 1 血壓分級方式[1]

血壓類別	收縮壓(mmHg)		舒張壓(mmHg)
正常	<120	及	<80
偏高	120-129	及	<80
高血壓第一期	130-139	或	80-90
高血壓第二期	>140	或	>90
高血壓危險期	>180	及/或	>120

4.樣本分布敘述：

表 2 樣本分布敘述

性別	血壓層級
男 23 人  (其中 5 人接受運動後量測)	正常 27 人，其中 0 人為運動後測量
	偏高 13 人，其中 4 人為運動後測量
	高血壓第一期 2 人，其中 2 人為運動後測量
女 21 人  (其中 3 人接受運動後量測)	高血壓第二期 6 人，其中 1 人為運動後測量
	高血壓危險期 0 人，其中 0 人為運動後測量

註：以總資料數計算，直行之間沒有對應關係。

### (三) 模型訓練及評估

本科展運用 CNN 的模型架構進行訓練，詳細步驟如下所述：

#### 1. 讀取時序頻譜圖資料，並利用檔名建立正確資料庫(Y)。

以建立不同血壓分級的資料庫為例，先從讀取的資料檔名中找到"D"，接著判斷其後面的數字為 0、1、2、3。

血壓分級矩陣的建立方式如下，若"D"後方數字為 0，則在 Y 中輸入[1,0,0,0]；若"D"後方數字為 1，則在 Y 中輸入[0,1,0,0]；若"D"後方數字為 2，則在 Y 中輸入[0,0,1,0]；若"D"後方數字為 3，則在 Y 中輸入[0,0,0,1]。

而在運動造成之高血壓及長期高血壓的分類中，[1,0]代表的是長期性的高血壓(0)，[0,1]代表的則是運動造成的高血壓(1)。

#### 2. 資料切割與打亂。

為避免模型在訓練時因資料順序產生偏見，我們以 NumPy 函數中的 `random.permutation` 將所有資料打亂。將資料打亂後，利用 `numpy` 函數中的 `split` 依據 6:2:2 分為 `train`、`validation`、`test` 三部分的資料。

考慮到此模型將有機會做為醫療用途，為了確定模型能夠真正判斷高血壓的資料，而非透過學習不同受測者之間的個人差異來判斷。為求更接近實際應用情形，我們也採取以人分類打亂資料的方式進行。

首先，利用 `random.sample` 從 D0、D1、D2、D3 資料中隨機抽取一筆資料，接著，讀取所抽出資料檔名上的受測者編號資訊，於後續程式讀入檔案時將此編號的檔案單獨存取至 `test` 資料的陣列中，並同樣使用 `random.permutation`，分別將 `test` 資料的陣列和剩餘資料的陣列打亂。最後，利用 `split`，將剩餘資料陣列依據 3:1 切割，分為 `train`、`validation` 兩部分。使用此方法打亂資料以後，模型在訓練時就不會遇到 `test` 資料庫中受測者的任何資料，自然也避免了模型是學習個人特徵而非血壓層級特徵的疑慮。

表 3 資料切割及其功能

	參與訓練與否	功能
train	是	進行模型訓練
validation	是	校正每輪 epoch 訓練後模型的準確性
test	否	預留作為判斷模型最後訓練結果的依據

### 3.建立模型

以 Keras 神經網路庫建立多層的 CNN 網路結構模型，每層結構則以一次 convolution 加一次 maxpooling，而為了避免 overfitting 的情況發生，我們使用了兩種解決方式，第一種是在 dense 層前加入 dropout，第二種是在卷基層間加入 Batch Normalization，再將 compile 函數將 loss 定義為'categorical\_crossentropy'、optimizer 定義為'adam'、metrics 定義為'accuracy'，並利用 ReduceLROnPlateau 使 learning rate 逐漸降低。

### 4.訓練模型

以 Keras 的 model.fit 訓練模型，並預測 validation 資料庫的結果，校正該次訓練。執行一次模型的訓練稱為一個 epoch，在本科展中，每次訓練執行 100 次 epoch 就能達到高正確率。

模型訓練的過程中，若 train 資料庫的正確解答和預測值的 loss，在訓練過程中已收斂，但 validation 資料庫和預測值的 val\_loss 持續發散，表示模型出現 overfitting，則需停止訓練，並重新調整模型。

若 validation 資料庫和預測值的差(val\_loss)也收斂，則在訓練結束後將 loss 最小的模型記錄下來，並以該模型對 test 資料庫進行預測，計算預測的正確率。

## 5.模型評估

為了能夠完整了解模型訓練狀況及是否有偏見，本科展利用混淆矩陣(Confusion Matrix)及相關效能衡量指標（準確率(Accuracy)、精確率(Precision)、召回率(Recall)、F1 score）進行模型評估。並以熱圖(Heatmap)將模型的判斷依據進行可視化，進一步了解模型判斷的情形。

以下為上述各項指標的介紹：

混淆矩陣(Confusion Matrix)：

此矩陣為用來評價模型結果的分析表，縱軸為實際標籤(Actual Class)，橫軸為模型預測結果(Predicted Class)。

圖 4 以二分模型（陽性(Positive, P)和陰性(Negative, N)）為例，若答案與預測皆為 P，則此結果計入 True Positives(TP)；若答案為 P 而預測為 N，則此結果計入 False Negative(FN)；若答案為 N 而預測為 P，則此結果計入 False Positive(FP)；若答案與預測皆為 N，則此結果計入 True negatives(TN)。

將模型判斷的情況填入圖表中後，便可利用 TP、FN、FP、TN 的數值進一步算出模型的準確率(Accuracy)、精確率(Precision)、召回率(Recall)及 F1 score。

		Predicted class	
		P	N
Actual Class	P	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

圖 4 confusion matrix 示意圖[2]



準確率(Accuracy)=  $(TP+TN)/(TP+FP+FN+TN)$  此項指標代表所有情況中判斷正確的比率

精確率(Precision)=  $TP/(TP+FP)$  此項指標代表在所有陽性樣本中，真正為陽性的比率

召回率(Recall)=  $TP/(TP+FN)$  此項指標代表事實為真時，有多少成功被判斷出來

F1 score=  $2 / ((1/ Precision) + (1/ Recall))$ ，此即為精確率與召回率的調和平均數。此項指標考量精確度和召回率，可作為綜合性的評量。

#### (四) 熱圖分析

熱圖(heatmap)：

熱圖分析若運用在 CNN 模型上，可以使模型層的判斷依據進行可視化處理。本科展運用的顏色配置為 jet，如下圖 5，模型在判斷時越偏重作為判斷依據的區塊，顏色會越偏紅色，反之，則顏色越偏藍[6]。



圖 5 heatmap 顏色指標

#### 伍、研究結果與討論

目前已蒐集 44 位受測者的血壓及心跳聲，其中有 21 位高血壓，進行資料處理後約有 900 個檔案。

本科展探討相關文獻資料後，認為深度學習能透過心音時序頻譜圖預測血壓層級，故嘗試以小資料庫對心音成分、模型架構、資料打亂方式分別進行了實驗設計、模型訓練及完整的分析。並嘗試以模型分辨運動造成的高血壓和長期性高血壓的差異。

一、以 CNN 模型判斷受測者的血壓層級，比對心音成分對模型訓練成果的影響

(一) 實驗目的：

由於 S2 和動脈瓣關閉造成的振動有關，其與心音的相關性應該比 S1 更高。我們希望藉由此實驗得出心音與高血壓在時序頻譜圖中是否有相關性，並比較以 S1+S2 的資料與僅 S2 的資料做模型訓練的成果，證明 S2 與高血壓的相關性更高。

(二) 模型架構與參數設定：

在第三與第四層間加入 dropout 以防止 overfitting，訓練資料集約 400 筆，訓練 70 epoch。

(三) 模型在不同資料庫下的訓練情況：

在此段落我們會以 loss 圖分別詳盡地描述 S1+S2 和 S2 的訓練過程

#### 1. 針對 S1+S2 資料做訓練

由下圖 7 的 loss 圖來看，training loss 在 1~2 epoch 急速下降，且於 3 epoch 後維持非常低的 loss；但 validation loss 在前 10 epoch 有小波動後整條線幾乎呈水平，且較 training loss 高了許多，故判斷其有輕微的 overfitting。

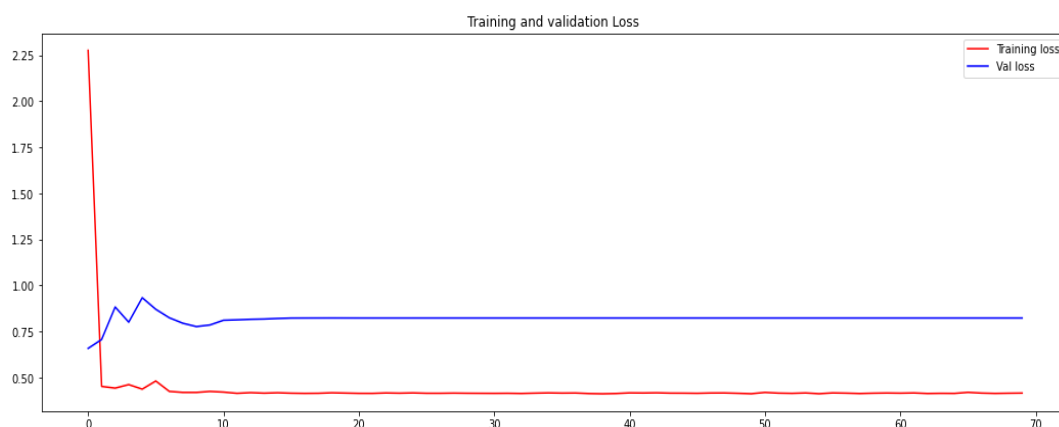


圖 6 針對 S1+S2 資料做訓練的 loss 圖

註：此圖的縱軸為 loss，橫軸為 epoch，且其後 loss 圖形皆如此

## 2. 僅有 S2 資料做訓練

由圖 8 可見 training loss 在 1~4 epoch 急速下降，且於 5 epoch 後維持非常低的 loss；而 validation loss 在 1~2 epoch 時有較大的波動，在 3~10 epoch 間波動持續減小，而從 10 epoch 開始 loss 圖幾乎保持水平線，代表模型已經收斂。綜合兩條線來看，training 和 validation 的 loss 皆快速收斂。相比於以 S1+S2 作為資料庫的訓練成果，validation loss 和 training loss 的差距較小，模型訓練結果較佳。

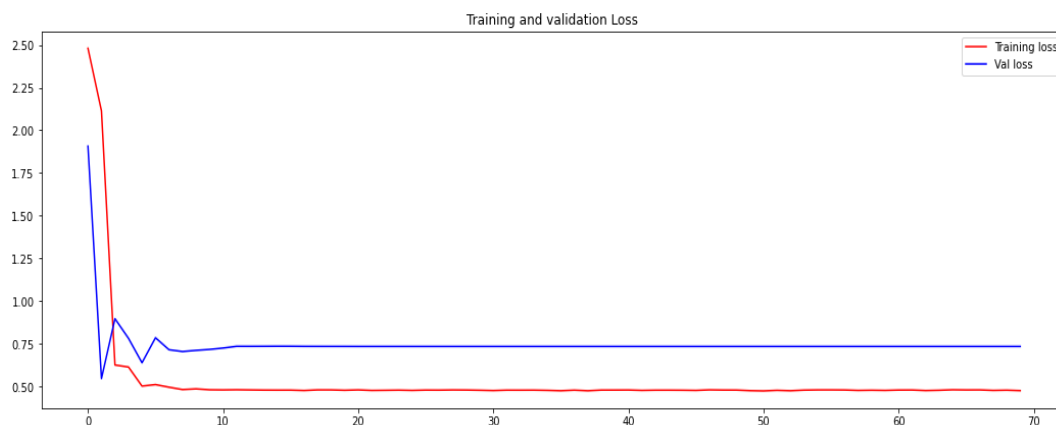


圖 7 針對僅 S1 資料做訓練的 loss 圖

### (四) 結果討論：

S2 作為資料庫相比於 S1+S2 作為資料庫 training loss 和 validation loss 的差距較小，模型收斂狀況較佳。以下針對 1. 模型各項指標評估數值、2. confusion matrix 和 3. heatmap 進行深入的分析及比較。

#### 1. 模型各項指標評估數值比較：

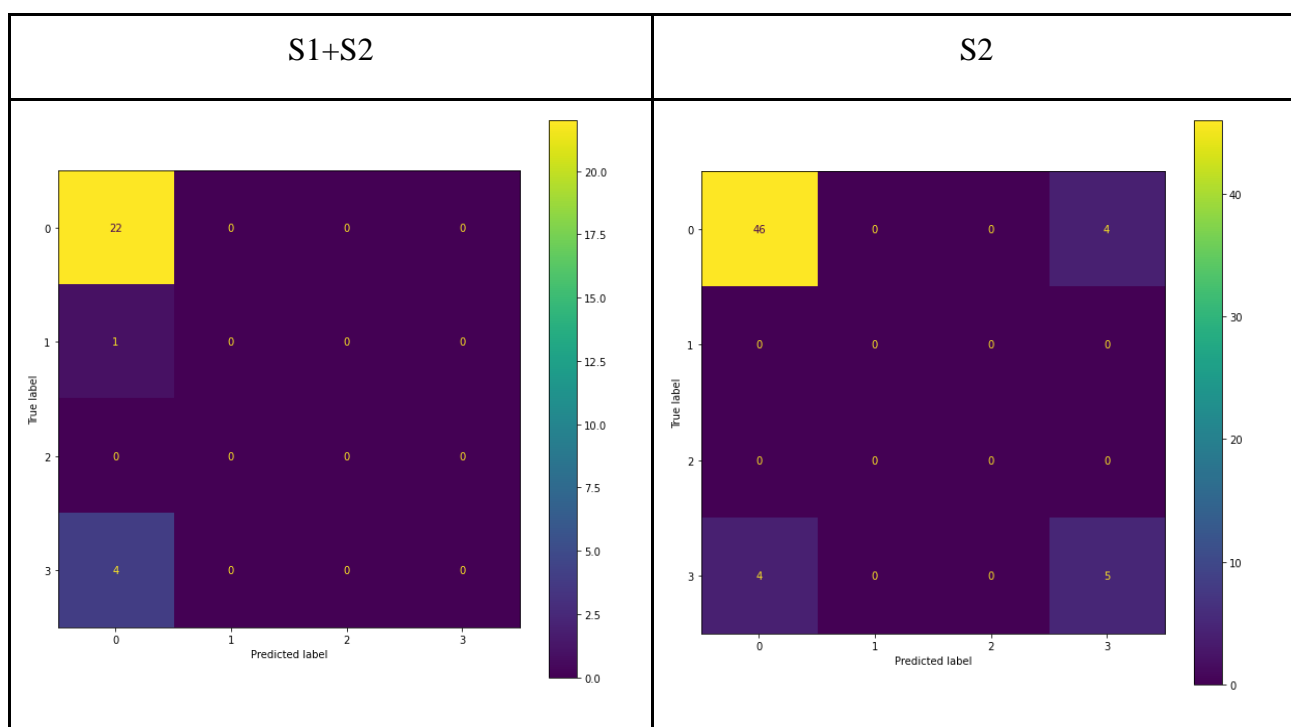
由表 4 可見，僅 S2 資料的訓練模型在各項指標表現中都更勝於 S1+S2 資料的訓練模型，由此可印證 S2 和高血壓更為相關的假設。

表 4 訓練不同資料庫的模型各項指標的比較

S1+S2	S2
Test Accuracy : 0.8148	Test Accuracy : 0.8889
Precision: 0.6639	Precision: 0.7901
Recall: 0.8148	Recall: 0.8889
F1 Score: 0.7317	F1 Score: 0.8366

### 3.Confusion Matrix 比較：

表 5 不同資料庫模型訓練結果的 confusion matrix 比較



如表 5 所示，訓練 S1+S2 資料的模型只答對了標籤為 0 的資料，且將所有資料預測為標籤 0；而訓練僅 S2 資料的模型，雖然在預測標籤 1 與標籤 3 時仍有容易混淆的跡象，但此模型仍有成功預測出非標籤 0 的資料。由以上分析可知，相較於 S1+S2 的資料，僅 S2 資料的訓練結果較佳。

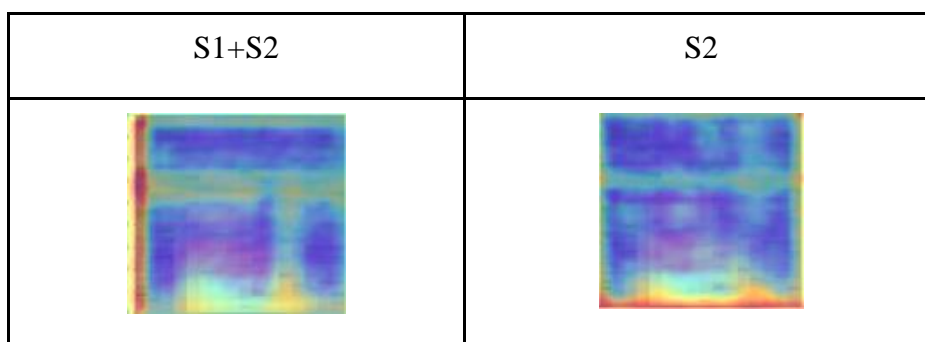
但由表 4 的評估數值來看，訓練 S1+S2 資料模型的各項數值指標仍維持一定的水準，我們推測其可能原因為資料的不平衡，test 中有相對多數的標籤 0 資料，造成模型即使將全部資料預測為 0，仍能成功預測多數為 0 的標籤。

另外，從表 5 中的 confusion matrix 會發現，S1+S2 的資料量較僅 S2 的資料少，是因為在早期的實驗過程中就有發現 S1+S2 資料訓練結果不如僅 S2 的資料，因此我們後續都只使用僅有 S2 的資料庫進行實驗，也只以 S2 的心音資料處理為主。

### 3. Heatmap 分析：

由下表 6 可看到於 heatmap 中看出模型相較於其他部分更重視 500Hz~700Hz 間與 300Hz 以下的區段，因此可以推測血壓反映在時序頻譜圖上的特徵是明顯的。

表 6 不同資料庫模型訓練結果的 heatmap 比較



### 4. 總結

受限於資料庫的太小且不平均，從 confusion matrix 中可看見模型有明顯偏見。然而就算在侷限的狀況下訓練，綜合上述評估指標和 heatmap 分析，我們仍然能得到重要結論：血壓的特徵能夠反映在時序頻譜圖上，並且 S2 在高血壓的判定上更為重要。

## 二、比對模型層數對模型訓練成果的影響

### (一) 實驗目的：

在進行本科展研究時，我們調整多項模型參數，甚至是模型架構，以求最佳訓練成果。在實驗二中，我們只提及結果差異較明顯的層數調整作為一項實驗。根據實驗一的結果，僅有 S2 資料的模型訓練結果較佳。然而相較於網路上大型資料庫，我們現有資料庫小非常多，因此我們想了解四層模型對我們的資料庫來說是否是最佳解。所以我們決定比較兩層、三層、四層和五層的模型訓練結果，並找出表現最佳的模型。

### (二) 模型架構及參數設定：

此實驗藉由調整模型的層數，找出表現最佳的結構。實驗涵蓋兩層、三層、四層和五層的模型架構。

### (三) 模型結構及訓練狀況：

以下會分別討論四種不同結構的模型在訓練過程的差異

#### 1.五層模型

##### (1) 模型結構：

```

Model: "sample_model"
-----
Layer (type)      Output Shape      Param #
-----
input_layer (InputLayer) [(None, 217, 223, 4)] 0

conv1 (Conv2D)      (None, 109, 112, 32) 1152

batch_normalization_5 (Batch Normalization) (None, 109, 112, 32) 128

activation_5 (Activation) (None, 109, 112, 32) 0

conv2 (Conv2D)      (None, 55, 56, 64) 18432

batch_normalization_6 (Batch Normalization) (None, 55, 56, 64) 256

activation_6 (Activation) (None, 55, 56, 64) 0

conv3 (Conv2D)      (None, 28, 28, 128) 73728

batch_normalization_7 (Batch Normalization) (None, 28, 28, 128) 512

activation_7 (Activation) (None, 28, 28, 128) 0

conv4 (Conv2D)      (None, 14, 14, 256) 294912

batch_normalization_8 (Batch Normalization) (None, 14, 14, 256) 1024

activation_8 (Activation) (None, 14, 14, 256) 0

conv5 (Conv2D)      (None, 7, 7, 512) 1179648

batch_normalization_9 (Batch Normalization) (None, 7, 7, 512) 2048

activation_9 (Activation) (None, 7, 7, 512) 0

global_average_pooling2d_1 (GlobalAveragePooling2D) (None, 512) 0

reshape_1 (Reshape) (None, 1, 512) 0

dropout_1 (Dropout) (None, 1, 512) 0

output_layer (Dense) (None, 1, 4) 2052
-----
Total params: 1,573,892
Trainable params: 1,571,908
Non-trainable params: 1,984
-----

```

圖 8 五層模型結構圖

## (2) 模型訓練 loss 圖與 accuracy 圖

下圖 9 中，training loss 在前 30 個 epoch 有幅度較明顯的下降，於 30 epoch 後下降幅度逐漸平緩；而 validation loss 在前 30 個 epoch 有大幅波動，且於 20~30 epoch 間最為明顯，而在 30~60 epoch 逐漸平緩，於 60 epoch 後幾乎呈水平線。綜合兩條線來看，雖然 validation loss 在最後仍比 training loss 高，但其值有收斂至約 0.3，而 training loss 也有收斂至近乎為 0 處。此外，兩線收斂趨勢皆非常明顯，故我們將此訓練視為成功的訓練。

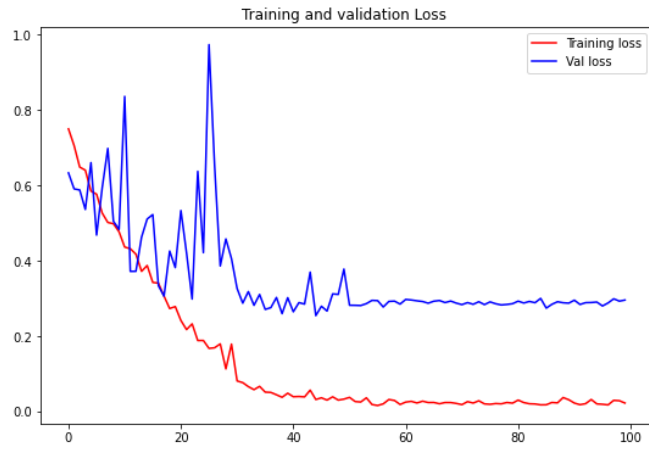


圖 9 以 S2 資料做四層模型訓練的 loss 圖

下圖 10 中，training accuracy 在前 40 個 epoch 持續漸漸上升，而於 40 epoch 後幾乎維持定值；而 validation accuracy 在前 40 個 epoch 有劇烈波動，尤其於 20~40 epoch 波動幅度最大，而於 30~60 epoch 波動逐漸平緩，於 60 epoch 其後波動幅度非常微小。綜合兩條線來看，雖然 validation accuracy 未像 training accuracy 一樣高，其準確率仍能達到約 90%，且 training accuracy 本身也達到約 98% 的高準確率，故此圖顯其訓練成果佳。

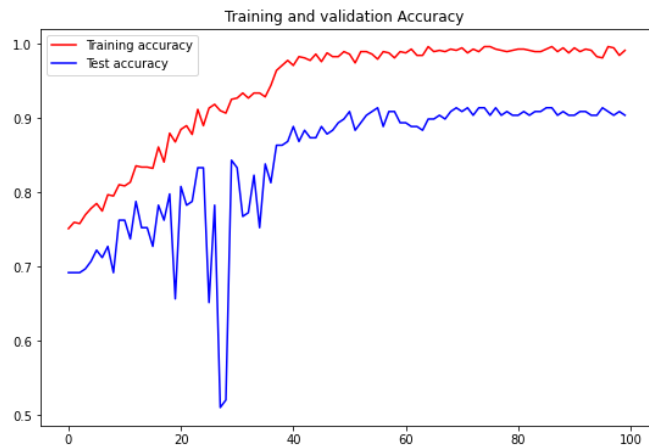


圖 10 以 S2 資料做四層模型訓練的 accuracy 圖



## 1. 四層模型

### (1) 模型結構：

Layer (type)	Output Shape	Param #
input_layer (InputLayer)	[(None, 217, 223, 4)]	0
conv1 (Conv2D)	(None, 109, 112, 32)	1152
batch_normalization_4 (Batch Normalization)	(None, 109, 112, 32)	128
activation_4 (Activation)	(None, 109, 112, 32)	0
conv2 (Conv2D)	(None, 55, 56, 64)	18432
batch_normalization_5 (Batch Normalization)	(None, 55, 56, 64)	256
activation_5 (Activation)	(None, 55, 56, 64)	0
conv3 (Conv2D)	(None, 28, 28, 128)	73728
batch_normalization_6 (Batch Normalization)	(None, 28, 28, 128)	512
activation_6 (Activation)	(None, 28, 28, 128)	0
conv4 (Conv2D)	(None, 14, 14, 256)	294912
batch_normalization_7 (Batch Normalization)	(None, 14, 14, 256)	1024
activation_7 (Activation)	(None, 14, 14, 256)	0
global_average_pooling2d_1 (GlobalAveragePooling2D)	(None, 256)	0
reshape_1 (Reshape)	(None, 1, 256)	0
dropout_1 (Dropout)	(None, 1, 256)	0
output_layer (Dense)	(None, 1, 4)	1028

=====  
Total params: 391,172  
Trainable params: 390,212  
Non-trainable params: 960  
=====

圖 11 四層模型架構圖

### (2) 模型訓練 loss 圖與 accuracy 圖

下圖 12 中，training loss 在前 40 個 epoch 有幅度較明顯的下降，而於 40 epoch 後下降幅度非常平緩；而 validation loss 在前 40 個 epoch 有大幅波動，尤其於 20~40epoch 間最為明顯，而在 40~60 epoch 波動逐漸平緩，於 60 epoch 後幾乎呈水平線。綜合兩條線來看，雖然 validation loss 在最後仍比 training loss 高出一些，但其值仍然有收斂至約 0.3，而 training loss 也有收斂至約 0.05 處。此外，兩線收斂趨勢皆非常明顯，故我們判斷這是一次成功的模型訓練。



圖 12 以 S2 資料做四層模型訓練的 loss 圖

下圖 13 中，training accuracy 在前 40 個 epoch 持續上升，而於 40 epoch 後幾乎維持定值；而 validation accuracy 在前 40 個 epoch 有劇烈波動，尤其於 20~30 epoch 波動最明顯，而於 30~60 epoch 波動逐漸平緩，於 60 epoch 其後波動幅度非常微小。綜合兩條線來看，雖然 validation accuracy 在最後未像 training accuracy 一樣高，其準確率仍能達到約 90%，且 training accuracy 本身也達到約 98% 的高準確率，故此圖顯示其訓練成果不錯。

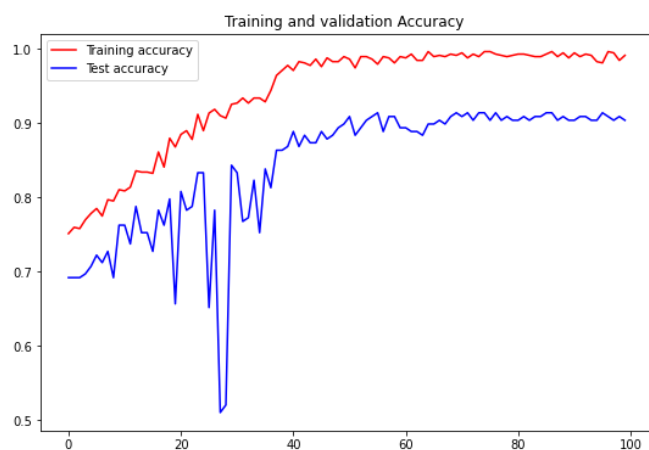


圖 13 以 S2 資料做四層模型訓練的 accuracy 圖

## 2.三層模型：

### (1)模型結構

```
Model: "sample_model"
Layer (type)      Output Shape      Param #
-----
input_layer (InputLayer)  [(None, 217, 223, 4)]  0
conv1 (Conv2D)      (None, 109, 112, 32)  1152
batch_normalization_3 (Batch Normalization) (None, 109, 112, 32)  128
activation_3 (Activation) (None, 109, 112, 32)  0
conv2 (Conv2D)      (None, 55, 56, 64)  18432
batch_normalization_4 (Batch Normalization) (None, 55, 56, 64)  256
activation_4 (Activation) (None, 55, 56, 64)  0
conv3 (Conv2D)      (None, 28, 28, 128)  73728
batch_normalization_5 (Batch Normalization) (None, 28, 28, 128)  512
activation_5 (Activation) (None, 28, 28, 128)  0
global_average_pooling2d_1 (GlobalAveragePooling2D) (None, 128)  0
reshape_1 (Reshape)  (None, 1, 128)  0
dropout_1 (Dropout)  (None, 1, 128)  0
output_layer (Dense)  (None, 1, 4)  516
-----
Total params: 94,724
Trainable params: 94,276
Non-trainable params: 448
```

圖 14 三層模型架構圖

### (2)模型訓練 loss 圖

下圖 15 中，training loss 在 1~20 epoch 有較明顯的下降趨勢，於 20 epoch 以後皆保持微小波動狀態；而 validation loss 則在 1~20epoch 有大幅波動，於 20~25 epoch 有較小的波動，其後皆為接近水平狀態。綜合兩條線來看，training loss 與 validation loss 都有明顯收斂情形，雖然 validation loss 最後並沒有收斂至緊貼 training loss，但兩線之值最終都達到約 0.5，此收斂狀況整體而言仍然不差。

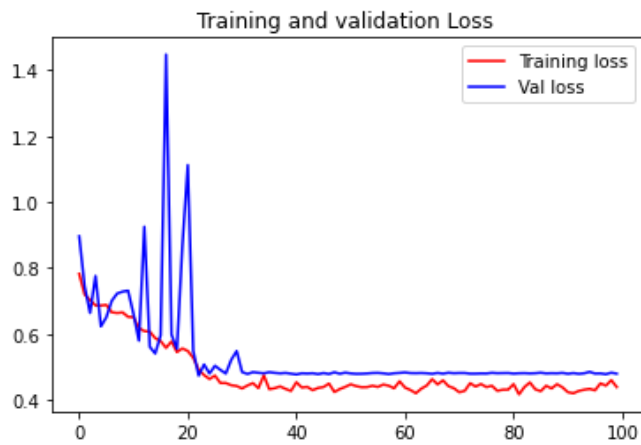


圖 15 以 S2 資料做三層模型訓練的 loss 圖

下圖 16 中，training accuracy 在 1~20epoch 有上升趨勢，於 20epoch 以後皆保持小幅度波動狀態；而 validation loss 則在 1~20epoch 有極大幅的波動，於 20epoch 後幾乎呈水平狀態。綜合兩條線來看，training accuracy 的準確率仍高於 validation accuracy，且 validation accuracy 並不像前者四層模型訓練的圖?一樣達到 90%的準確率，其準確率僅有約 80%，training accuracy 也沒有達到約 98%的高準確率，可見其模型訓練並未如同四層模型一樣好。

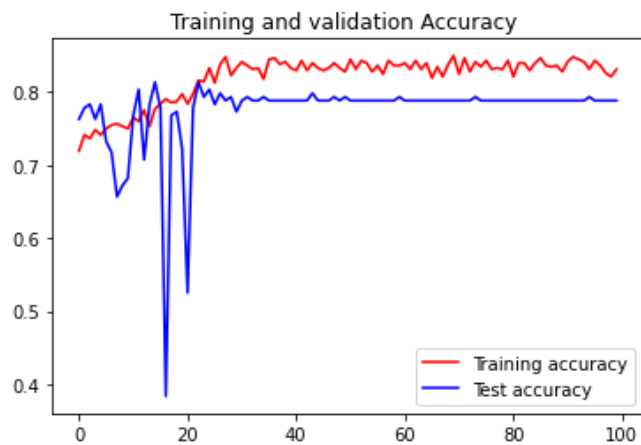


圖 16 以 S2 資料做三層模型訓練的 accuracy 圖

### 3.兩層模型

#### (1)模型結構

```
Model: "sample_model"
```

Layer (type)	Output Shape	Param #
input_layer (InputLayer)	[(None, 217, 223, 4)]	0
conv1 (Conv2D)	(None, 109, 112, 32)	1152
batch_normalization_10 (Batch Normalization)	(None, 109, 112, 32)	128
activation_10 (Activation)	(None, 109, 112, 32)	0
conv2 (Conv2D)	(None, 55, 56, 64)	18432
batch_normalization_11 (Batch Normalization)	(None, 55, 56, 64)	256
activation_11 (Activation)	(None, 55, 56, 64)	0
global_average_pooling2d_4 (Global Average Pooling2D)	(None, 64)	0
reshape_4 (Reshape)	(None, 1, 64)	0
dropout_3 (Dropout)	(None, 1, 64)	0
output_layer (Dense)	(None, 1, 4)	260

```
=====  
Total params: 20,228  
Trainable params: 20,036  
Non-trainable params: 192  
=====
```

圖 17 兩層模型架構圖

#### (2)模型訓練 loss 圖

下圖 18 中，**training loss** 在 1~20 epoch 有些微下降，20 epoch 後皆維持小波動的平衡狀態；而 **validation loss** 則在 1~20 epoch 間有大幅波動情形，20~40 epoch 間有些微波動，而 40 epoch 後皆維持水平狀態。綜合來看，**training loss** 與 **validation loss** 皆有收斂，雖然訓練最後 **validation loss** 仍然比 **training loss** 高出了一些，其值約為 0.7，而 **training loss** 也僅達約 0.65，可見其整體雖然有收斂，但狀況並不是太好。



圖 18 以 S2 資料做兩層模型訓練的 loss 圖

下圖 19 中，training accuracy 一直都呈現極小幅度的波動，甚至趨近於水平；而 validation loss 則在 1~20 epoch 間有極大的波動，於約 25 epoch 有些微波動，其後皆維持水平狀態。綜合來看，training accuracy 大約僅達到 75% 的準確率，而 validation loss 也僅約達到 70% 的準確率，可以明顯的看出其訓練效果較四層與三層模型差。

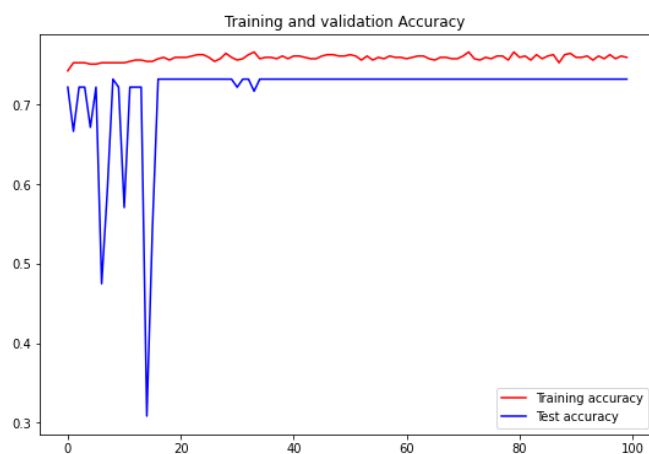


圖 19 以 S2 資料做兩層模型訓練的 accuracy 圖

#### (四) 結果討論：

由以上的實驗結果來看，雖然四層模型訓練狀況不突出，其訓練出來的結果數值評估仍然最佳，其準確率可達 0.9205，因此我們推斷四層模型相對來說是現階段最適合我們資料庫的模型，也是最有潛力的模型。

1. 模型各項指標評估數值比較：

表 7 四層模型、三層模型與兩層模型的各項指標比較

兩層模型	三層模型	四層模型	五層模型
Test Accuracy : 0.7841	Test Accuracy: 0.8409	Test Accuracy:0.9432	Test Accuracy: 0.8977
Precision : 0.8091	Precision: 0.8466	Precision: 0.9432	Precision: 0.8929
Recall : 0.7841	Recall: 0.8409	Recall: 0.9432	Recall: 0.8977
F1 score : 0.7079	F1 score: 0.8199	F1 score:0.9427	F1 score: 0.8937

由表上 7 可知，四層模型的各項指標皆較兩層、三層與五層的模型佳。我們推測雖然目前資料庫並不像網路上的資料庫一樣龐大，但已經適合相對較深的模型，然而五層的模型又對現有資料庫來說太深。其次，兩層與三層模型因為層數較少，無法萃取出較細微的特徵，因此訓練結果相對較差，而由下圖 20 也可明顯看出四層模型在各項指標皆表現得最好。

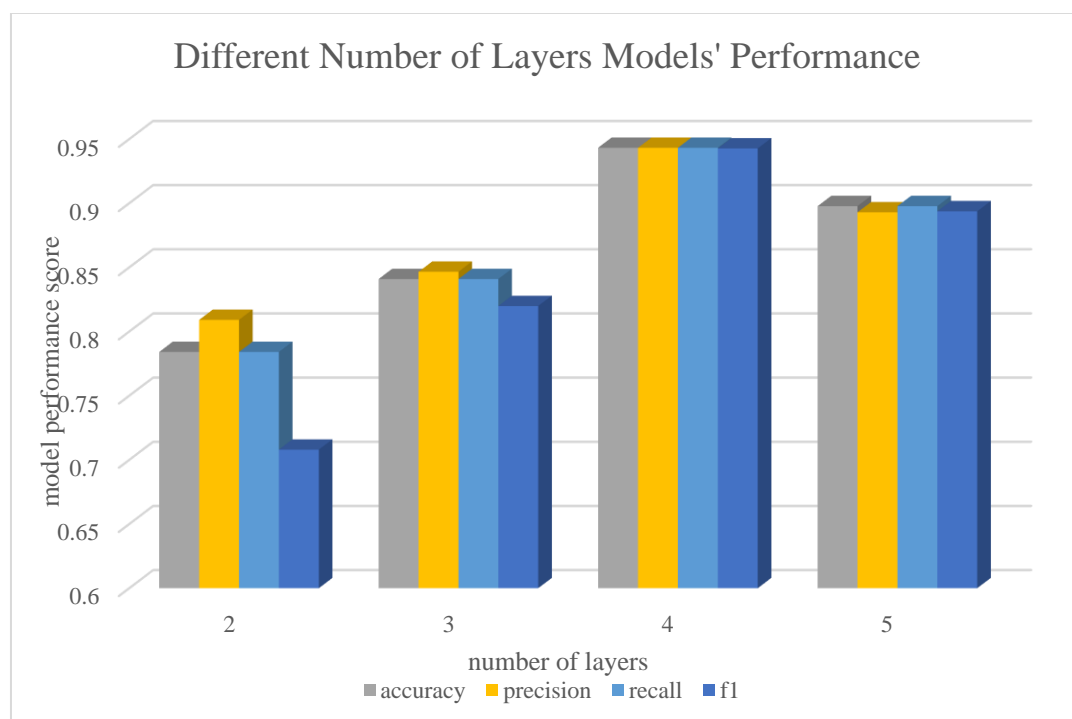
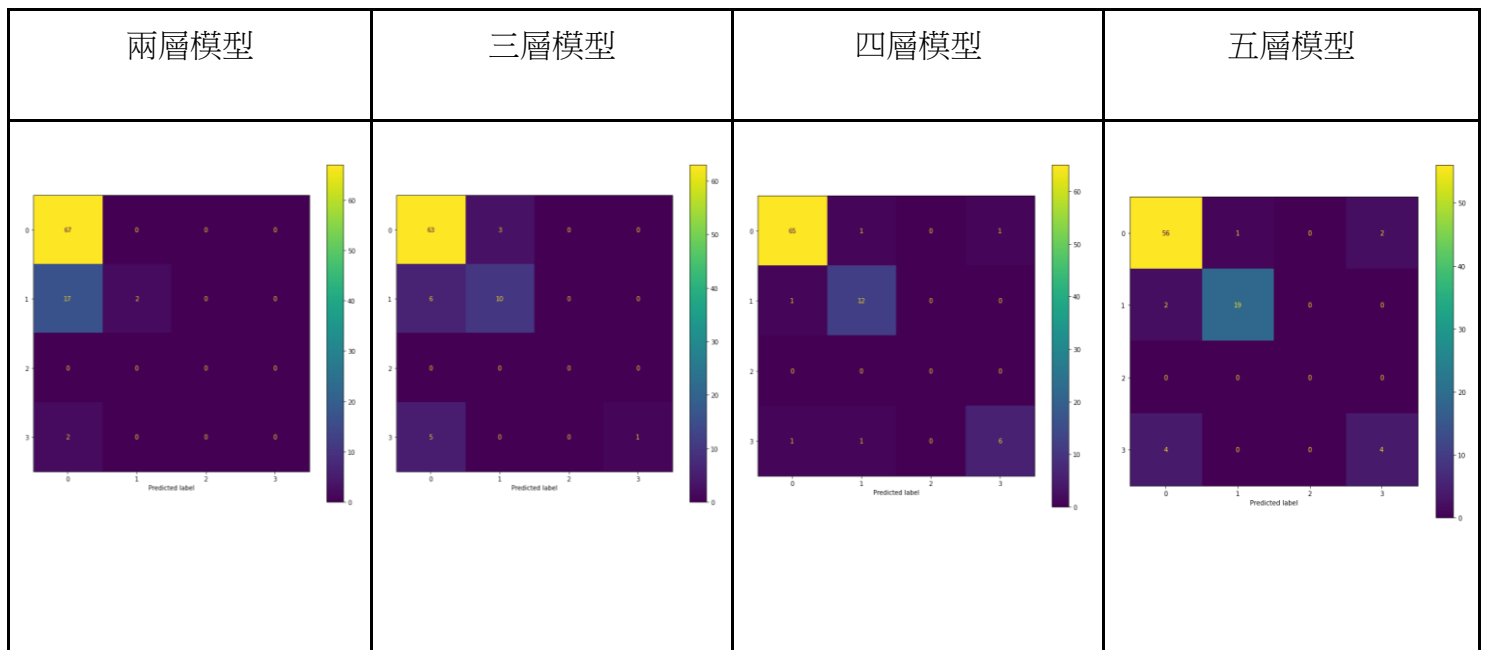


圖 20 不同層數模型的訓練結果圖表

## 2. Confusion Matrix 比較：

由表 8 可知，兩層模型在預測上具有嚴重偏見，幾乎將所有資料皆預測為標籤 0；而三層模型有將較多資料成功預測出標籤 1 與標籤 3，但在標籤 1 與標籤 3 處仍有較明顯的混淆現象；四層模型於 88 筆測資中僅有兩筆標籤 3、一筆標籤 1 和兩筆標籤 0 答錯，可知此模型擁有較少混淆的狀況，訓練結果最佳；五層模型則在標籤 3 的判讀上混淆情形較嚴重。

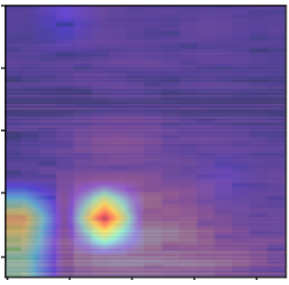
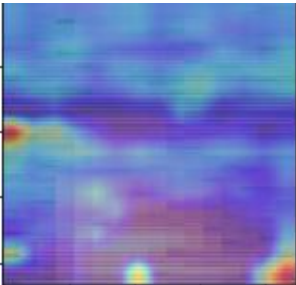
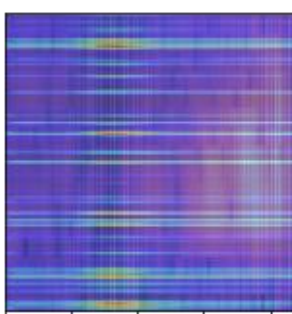
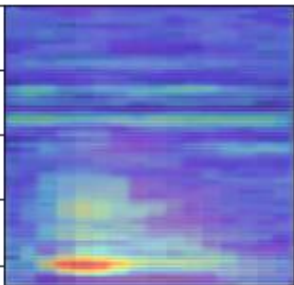
表 8 四層模型、三層模型與兩層模型的 confusion matrix 比較





### 3. Heatmap 分析：

表 9 四層模型、三層模型與兩層模型的 heatmap 比較

	五層模型	四層模型	三層模型	兩層模型
heatmap				
頻率分布	窄	廣	廣	廣，但有較明顯的判斷區域
集中程度	集中	整體分散，但有特別集中的區域	較稀疏	還算集中
形狀	圓形	橫向帶狀	橫向細絲狀	橫向分布為主，較低頻處有梯形分布

### 4. 總結

在此項實驗中，我們發現四層模型的訓練成果最佳，兩層模型最差而三層模型與五層模型居中。然而模型層數的選擇和資料庫大小、標籤方式和資料品質皆有關係，因此日後擴充資料庫後，在資料層數的選擇上有可能和現狀不同。

### 三、比對受測者心音差異對模型訓練成果的影響

#### (一) 實驗目的：

在先前的研究中，我們將資料打亂的方式皆為所有資料隨機打亂。這會造成模型在訓練時可能已經遇過 test 資料庫中的受測者，並學習此受測者相關的心音特徵，而非學習血壓層級的心音特徵。在用 test 資料庫評估模型時，模型有可能是預測出受測者，再根據先前訓練結果「背」出其血壓層級。為避免此情況，我們在本實驗中採取不同的資料打亂方式，也就是於肆－三－(三)－3 資料切割與打亂中所提到的第二種方式，以了解模型對於利用心音時序頻譜圖判斷血壓層級更真實的準確性，也更貼近現實應用。

#### (二) 模型架構及參數設定：

模型架構同圖 8 所示，訓練 100 epoch。

#### (三) 模型採用新的資料打亂方式後的訓練過程：

##### 1. 隨機選取 D0 資料中的一人作為 test data

下圖 21 中，training loss 在前 20 個 epoch 有較明顯的下降，而於 40 epoch 後便趨近於平緩；而 validation loss 在前 20 個 epoch 有大幅波動，其於 20epoch 後逐漸趨於水平。綜合兩條線來看，雖然 validation loss 在最後仍比 training loss 高出一些，但其值仍維持在約 0.3~0.4 處，而 training loss 也達到約 0.1~0.2 處，且兩線皆可明顯看出收斂趨勢，故此訓練成果極好。

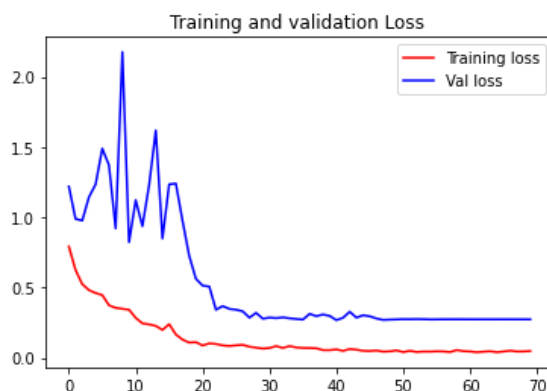


圖 21 新的資料打亂方法中以 D0 的其中一人作為 test 資料的訓練 loss 圖

下圖 22 中，training accuracy 在前 20 個 epoch 有較明顯的上升，而於 40 epoch 後便趨近於水平；而 validation accuracy 在前 20 個 epoch 有大幅波動，於 20epoch 後逐漸趨於水平。綜合兩條線來看，雖然 validation accuracy 在最後達到約 85%，雖然不如 training loss 達到將近 100%，但其訓練結果仍算佳。

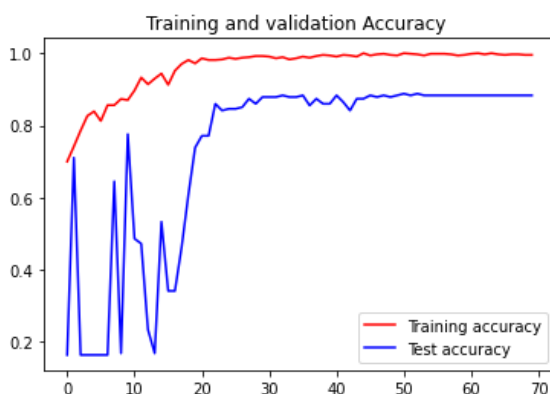


圖 22 新的資料打亂方法中以 D0 的其中一人作為 test 資料的訓練 accuracy 圖

## 2. 隨機選取 D1 資料中的一人作為 test data

下圖 23 中，training loss 在前 20 個 epoch 有較明顯的下降，於 20~40 epoch 下降幅度逐漸平緩，在 40epoch 後只剩小波動；而 validation loss 在前 20 個 epoch 有較大波動，其後逐漸平緩。綜合兩條線來看，雖然 validation loss 在最後仍比 training loss 高出一些，但其值仍維持在約 0.4 處，而 training loss 也達到約 0.15 處，且兩線皆有收斂趨勢，故此可算是有效的訓練。

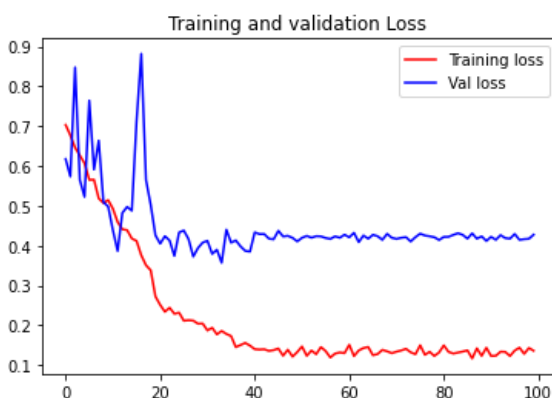


圖 23 新的資料打亂方法中以 D1 的其中一人作為 test 資料的訓練 loss 圖

下圖 24 中，training accuracy 在前 40 個 epoch 持續上升，於 40 epoch 逐漸平緩；而 validation accuracy 在前 20 個 epoch 波動較大，其後逐漸平緩。綜合兩條線來看，雖然 validation accuracy 在最後僅達約 90%，仍比 training accuracy 的約 98% 低了一些，但其準確率仍有達一定的水準。



圖 24 新的資料打亂方法中以 D1 的其中一人作為 test 資料的訓練 accuracy 圖

### 3. 隨機選取 D3 資料中的一人作為 test data

下圖 25 中，training loss 值維持緩慢的下降；而 validation loss 於前 20 個 epoch 有較明顯的波動，從 20 epoch 開始波動幅度減小且緩緩下降，至 80~100epoch 時 波動有變大的趨勢。綜合兩線來看，其收斂狀況皆很明顯，而至 80~100epoch 時皆仍有明顯波動，可嘗試增加訓練的 epoch 數，也許能得到更低得 loss 值。



圖 25 新的資料打亂方法中以 D3 的其中一人作為 test 資料的訓練 loss 圖

下圖 26 中，training accuracy 值維持緩慢的上升；而 validation loss 於前 20 個 epoch 有較大幅度的波動，從 20 epoch 開始波動減緩且緩慢上升。綜合兩線來看，其 accuracy 皆達約 80%，雖然不是很高，但因為訓練至最後不論 training accuracy 或 validation accuracy 都有相對大的波動，顯示它應該還能夠繼續訓練得到更好的結果。

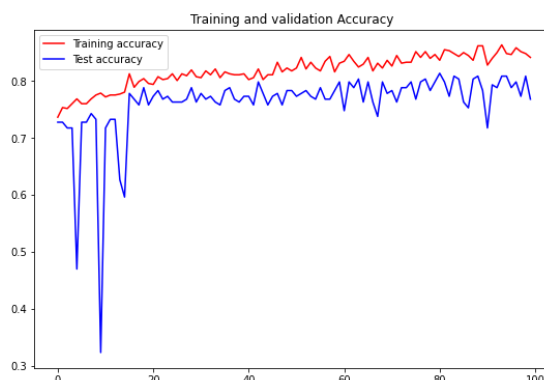


圖 26 新的資料打亂方法中以 D3 的其中一人作為 test 資料的訓練 accuracy 圖

#### (四) 結果討論：

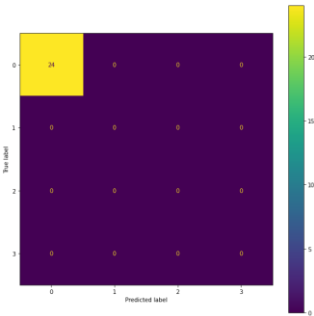
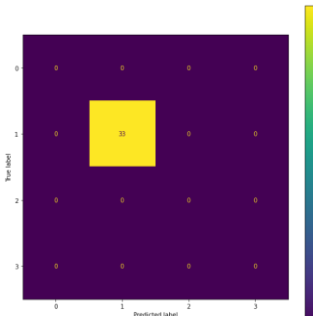
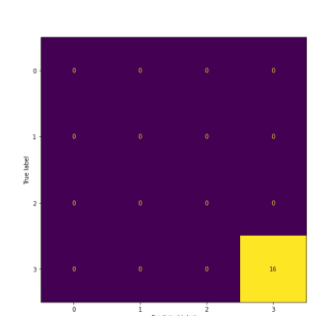
##### 1. 模型各項指標評估數值比較

由表 10 可知，在利用新的資料打亂方法後模型的準確率大大提升，不論以 D0、D1 或 D3 都能夠全部答對，是出乎意料的結果。此結果代表模型能夠辨識從來沒接觸過的人的資料，在避免了模型認人的特徵而非高血壓的特徵的「作弊」方法後，模型準確率不僅沒有下降，反而擁有跳躍性的提升，此結果證明了此方法在實際運用上的可行性，也再次證明了血壓與心音間是有關連性的。

表 10 以不同資料打亂方式訓練模型的各項指標比較

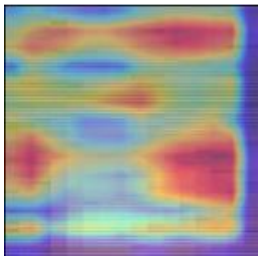
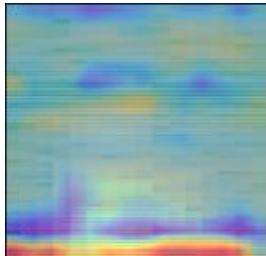
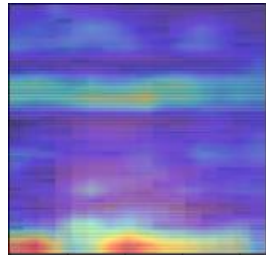
	D0	D1	D3
Test Accuracy :	1.0	1.0	1.0
Precision :	1.0	1.0	1.0
Recall :	1.0	1.0	1.0
F1 score :	1.0	1.0	1.0

		
--	--	---

### 3.Heatmap 分析

表 11 以不同資料打亂方式訓練模型的 heatmap 比較

	D0	D1	D3
			
頻率分布	廣，幾乎涵蓋所有範圍	廣，幾乎涵蓋所有範圍	窄，僅低頻處與中高頻處有分布
集中程度	較分散，但在特定部分有較明顯得集中區域	較分散，但低頻處有較集中的頻率帶	較集中，主要集中在低頻處與中高頻處
形狀	主要為橫向帶狀，亦有少數塊狀分布	主要為帶狀分布	帶狀分布

從上表 11 可以看到，以受測者為依據打亂後進行模型訓練所產生的 heatmap 皆為橫向分布而非垂直或斜向分布，因此可以推斷應該存在能使模型辨認出血壓高低的特定頻率，才會顯示出橫向的條紋。且相較於所有資料的打亂方式，模型判斷的分布區域更廣，我們推測可能是因為相較於原本的資料打亂方式，新的資料打亂方式排除了用人的特徵辨認的可能性，因此，若要正確的判斷出答案，可能就需要更多且更細部的特徵，而此情況也逼迫模型去學習那些特徵，也就連帶使其準確率提升。

#### 4.總結

從以上的訓練結果來看，改變資料打亂方式後準確率不僅沒有下降，還變高了，此結果也進一步證明了心音與血壓有相關性，也顯示出此技術的可行性與實際應用的可能性，且由此我們可以推斷出：

- (1)人的生理差異在心音時序頻譜圖上有對應的特徵。
- (2)不同血壓層級在心音時序頻譜圖上會有不同的頻率成分特徵，因此排除人的特徵後有相當好的訓練結果。

#### 四、不同頻率區段與高血壓相關程度比較

##### (一) 實驗目的：

從前面的實驗中我們可以從 heatmap 中看出模型較重視的部分，為了進一步驗證哪部分的頻率區段在模型判斷時擁有較重要的特徵，因此我們規劃了此實驗，藉由切割心音的不同頻率區段作為資料庫訓練模型，以模型表現，觀察哪個頻率區段訓練出的模型表現較好，藉此判斷其為較重要的頻率區段。

##### (二) 模型架構及參數設定：

模型架構如同圖 8，訓練 100 個 epoch。



(三) 模型訓練過程及結果

表 12 以不同頻率區段訓練模型所得到的 loss 圖、accuracy 圖及各項指標的結果數值整理

	loss 圖	accuracy 圖	結果
0~200			loss: 0.5056 accuracy: 0.8182 precision: 0.7460 recall: 0.8182 f1: 0.7772
200~400			loss: 0.6635 accuracy: 0.7614 precision: 0.6799 recall: 0.7614 f1: 0.7120
400~600			loss: 0.6917 accuracy: 0.750 precision: 0.7130 recall: 0.750 f1: 0.7189
600~800			loss: 0.7022 accuracy: 0.7386 precision: 0.6659 recall: 0.7386 f1: 0.6710
800~1000			loss: 0.8169 accuracy: 0.6477 precision: 0.5515 recall: 0.6477 f1: 0.5848

(四) 結果討論：

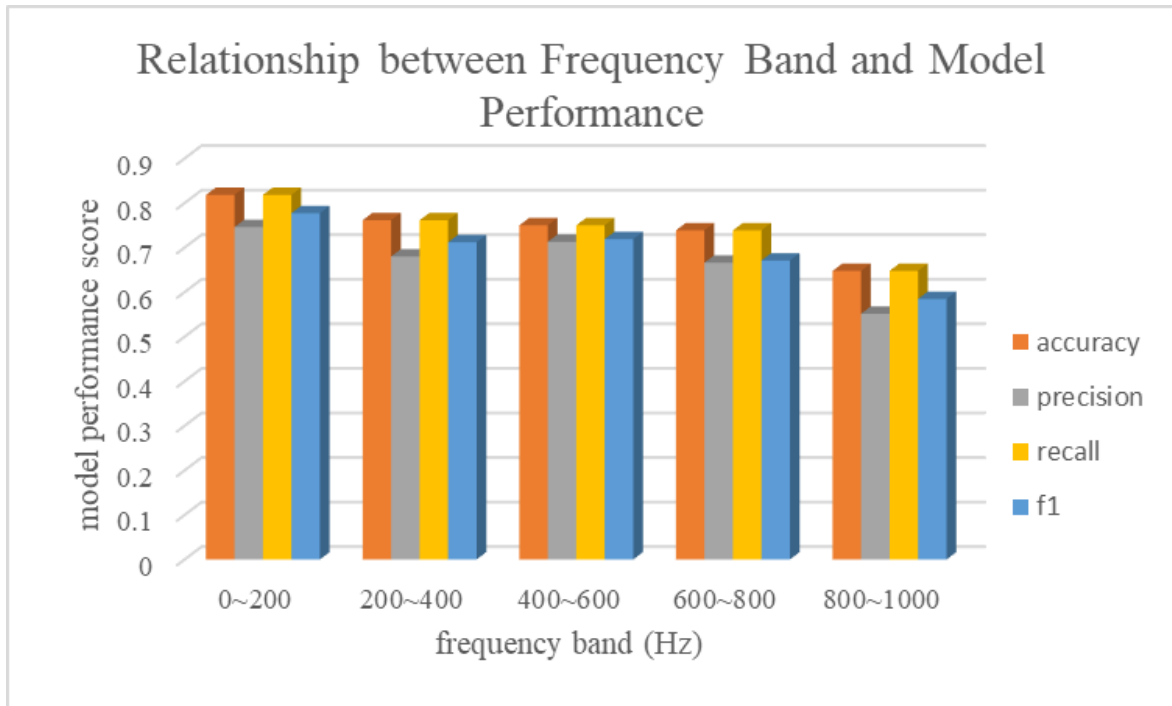


圖 27 以不同頻率區段的資料訓練模型的結果圖

從上圖 27 可看出越低頻率段所訓練出的模型準確率越高，而越高頻率段所訓練出的模型準確率越低的趨勢，而此圖中又可看出在 400~600Hz 的頻率段的表現不同於原本應有的下降趨勢，反而稍微比 200~400Hz 頻率段所訓練出的模型表現佳，因此我們猜測 0~200Hz 以及 400~600Hz 可能為模型在判斷上的關鍵之處。

五、建立模型辨別長期高血壓與運動後高血壓心音之差異

(一) 實驗目的：

運動造成的高血壓和長期高血壓成因並不相同，我們希望能透過此實驗從心音時序頻譜圖觀察到兩者差異，將來也可作為相關學術研究的基礎。另外，模型若能夠精準分類運動造成之高血壓和長期高血壓的能力，即可避免模型將運動所造成的正常生理反應誤判為長期高血壓。

## (二) 模型架構及參數設定：

此項實驗訓練資料集共 162 筆，其中 68 筆是長期性的高血壓，94 筆是運動造成的高血壓。

我們從實驗二的結果得知，兩層卷積層的模型訓練結果較佳，故在此實驗中皆使用兩層卷積層的模型進行訓練。

## (三) 模型訓練過程

可以從圖 28 中看到 training loss 與 validation loss 在 1~3 epoch 時皆急速下降，且於 4~70 epoch 仍持續下降，兩者 loss 值皆有收斂，且擬和狀態極佳，應屬於有效的模型訓練。

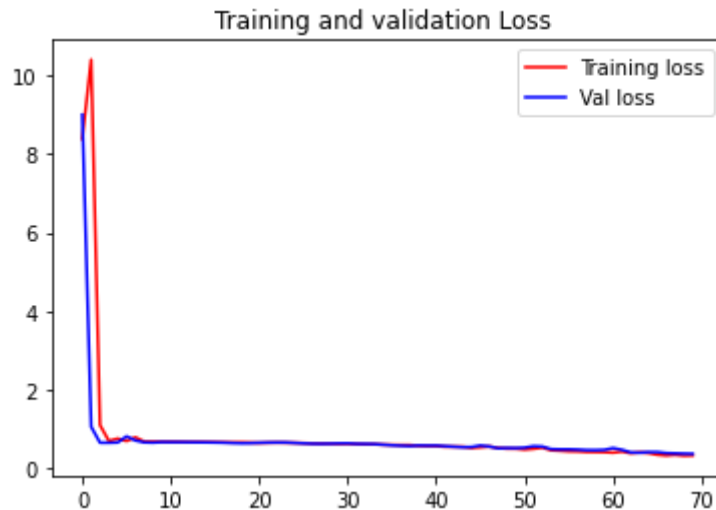


圖 28 區分運動後高血壓與長期高血壓模型訓練的 loss 圖

## (四) 結果討論：

### 1. 模型各項評估指標數值

Test Accuracy : 0.9394

Precision: 0.9449

Recall: 0.9394

F1 score: 0.9383

## 2. 模型預測的 confusion matrix

以圖 29 的 confusion matrix 的結果來看，模型的預測狀況非常準確，僅有 2 筆資料預測錯誤，訓練成果極佳。

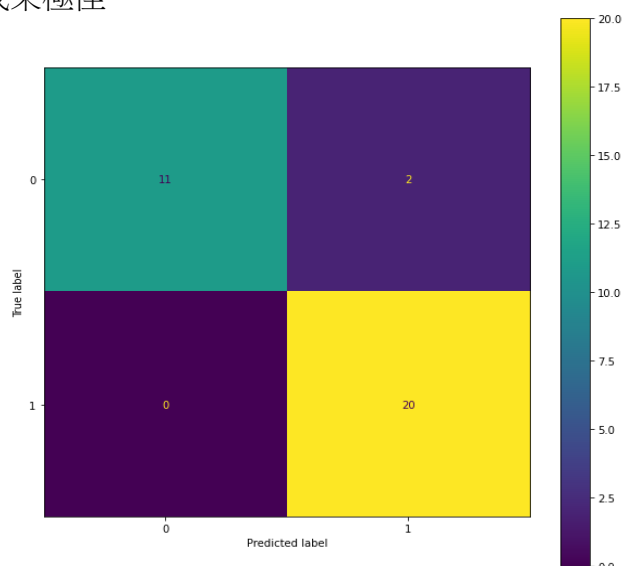


圖 29 區分運動後高血壓與長期高血壓模型訓練的 confusion matrix

## 3. 總結

目前本實驗的資料集僅約 160 筆資料但綜合各項指標可見模型訓練結果相當好。模型在訓練過程中有 33 筆資料沒有接觸過，但其中只有 2 筆是錯誤的。雖然目前資料庫較小，但已經證實運動造成的高血壓和長期性高血壓可以透過 CNN 對心跳聲的頻譜圖進行預測來分辨。未來會蒐集更多資料，並在此方向上進行更深入的研究。

## 陸、結論與應用

### 一、高血壓和心音有高度關聯

由上述結果可知，高血壓和心音有高度關聯性。以心跳聲的頻譜圖訓練深度學習模型後預測血壓層級，經過實驗一和實驗二，我們發現訓練最好的方式是使用僅有 S2 資料的四層模型，其準確率可達 0.9205。另外，將資料庫改為以人為基準打亂後，排除模型學習到人的心音特徵而非血壓之心音特徵的可能性，而其準確率不但沒有降低，反而在分別以 D0、D1、D3 作為 test data 時升高至 100%。從以上實驗結果可知，血壓層級在心音之時序頻譜圖應有對應的頻率成分特徵。

## 二、心音的 0~200Hz 與 400Hz~600Hz 頻率段可能與高血壓具有較高的關聯性

透過 heatmap，我們可以將模型在訓練時重視的特徵透過顏色標記出來。將 heatmap 疊加上座標軸後，我們發現多數 heatmap 都有將約在 0Hz~200Hz 與 500Hz~600Hz 的部分標記為相對重要的判斷依據，且從實驗四：不同頻率區段與高血壓相關程度比較的實驗中可以得出相同的結論，因此我們猜測心音中的 0Hz~200Hz 與 500Hz~600Hz 區段可能為血壓影響心音的主要部分。

## 三、心血管的結構差異在心音之時序頻譜圖上有對應之特徵

長期的高血壓會破壞心血管系統，使心臟和血管發生結構性改變，也因此大大提升罹患心臟病和相關併發症的機率。相較之下，運動造成的高血壓是為了滿足肌肉對氧氣的需求而造成的短期性血壓升高。從實驗五：建立模型辨別長期高血壓和運動後高血壓心音之差異可以發現模型可以分辨出長期的高血壓和運動造成的高血壓，此結果也意味著模型不僅能分辨出血壓層級，甚至能分辨出同為高血壓，但有結構性差異的兩類資料。我們可以知道，心臟和血管的結構應在心音之時序頻譜圖上有相對應之特徵，並且我們的模型透過訓練也可分辨出此特徵。

## 四、此研究的商業化應用

測量血壓時，一般人都是使用電子血壓機來量測，然而此方法無法持續性監測，也不適合手臂有傷的人使用。此外，市面上目前可測量血壓的穿戴型裝置，主要是利用光學測量血壓「變化」，必須先輸入原始血壓作為基準，每隔一段時間也需重新校正。相較之下，本研究所提出的方法是利用心跳聲直接判斷血壓層級，不須進行校正也可連續監測、追蹤。不管是和傳統的血壓計或現代性智慧穿戴型裝置比較，本研究在方便性和準確性上都是更勝一籌的。

## 五、未來展望

從實驗五中，我們驗證了結構上改變的特徵可以顯示於時序頻譜圖上。並且模型可以透過學習判斷出此特徵。這項重要結論表示會使心臟和血管的結構改變的疾病都有機會利用本研究提出的方法判斷，有廣泛的應用潛力。在技術方面，未來可以嘗試使用生成式模型，透過對抗式學習也許能進一步提升模型準確率。在應用方面，未來若本研究和硬體設備結合，有望製造出能方便、便宜，且能持續監測血壓的穿戴式裝置

## 柒、研究遇到的困難及解決方式

### 一、心音響度低，蒐集過程易被噪音影響

在剛開始蒐集資料時，我們使用手機貼著衣服錄製心跳聲，然而錄到的大多是衣服的摩擦聲而非心跳聲。因此，我們購買了醫用聽診器，並嘗試以手機直接抵在耳竇上，仍然沒有成功。經嘗試後決定將聽診器 Y 管以上的部分都剪掉，並將收音麥克風直接塞進聽診器導管內，錄音品質也因此有了顯著提升。



圖 30 聽診器連接麥克風的裝置圖

## 二、心音擷取不在正確區段

剛開始進行研究時，我們曾嘗試以程式進行自動切檔：在偵測到超過一定音量的波後回推一段固定時間做為開始點，再向後切一段固定時長的波。然而每個受測者的心音長度並不相同，也導致程式運作上經常出現錯誤，也沒有切在正確的心音區段。因此我們改使用 Audacity 除噪和切檔。除噪時，我們是先選取我們要濾除的噪音特徵再進行噪音抑制，以防意外刪除了重要的心音特徵。我們在查找文獻時，也發現有相關研究提出以信號處理方式提取 S2，並以 S2 進一步分為主動脈瓣關閉音和肺動脈瓣關閉音，類似技術將來也可作為我們資料處理的一環[3]。

## 三、模型評估方式的轉變

最初進行研究時，僅用最單純的準確率作為判斷模型的依據，但模型訓練的狀況和各項標籤誤判的比率都無從得知。因此我們引入了 loss 和 accuracy 隨 epoch 變化的圖和 confusion matrix 來評估，並對訓練中的 overfitting 或模型的偏見進行對應的處理，研究方向也因此變得更有條理。

## 捌、參考文獻

- [1] American Heart Association.(2022) .Understanding Blood Pressure Readings.  
Retrieved from American Heart Association Web site:  
<https://www.heart.org/en/health-topics/high-blood-pressure/understanding-blood-pressure-readings>
- [2] Eugenia Anello. (2021, February 23). How to evaluate your model using the Confusion Matrix. Medium. Retrieved from  
[http://rasbt.github.io/mlxtend/user\\_guide/evaluate/confusion\\_matrix/](http://rasbt.github.io/mlxtend/user_guide/evaluate/confusion_matrix/)
- [3] Jingping Xu; L.-G. Durand; P. Pibarot. (2001, March). Extraction of the aortic and pulmonary components of the second heart sound using a nonlinear transient chirp signal model. IEEE Transactions on Biomedical Engineering, 46(3), 277-283. doi: 10.1109/10.914790. Retrieved  
from <https://ieeexplore.ieee.org/document/914790>
- [4] Kenji Nakamura, Takeshi Fujiwara, Satoshi Hoshide, Yusuke Ishiyama, Mizuri Taki, Seiji Ozawa, Kazuomi Kario. (2021, April 23). Differences in exercise-induced

blood pressure changes between young trained and untrained individuals.

PubMed. doi: 10.1111/jch.14177. <https://pubmed.ncbi.nlm.nih.gov/33455054/>

- [5] Linha (Lina) M. Mohammed, Meera Dhavale, Mohamed K. Abdelaal, A B M Nasibul Alam, Tatjana Blazin, Dhruvil Prajapati, Jihan A. Mostafa. (2020, December 9). Exercise-Induced Hypertension in Healthy Individuals and Athletes: Is it an Alarming Sign? Cureus. doi: 10.7759/cureus.11988  
<https://www.cureus.com/articles/42807-exercise-induced-hypertension-in-healthy-individuals-and-athletes-is-it-an-alarming-sign>
- [6] MathWorks.(2022).Jet. Retrieved from American Heart Association Web site:  
<https://www.mathworks.com/help/matlab/ref/jet.html>
- [7]Sumit Saha. (2018, December 16). A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way. Medium. Retrieved from  
<https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>
- [8] World Health Organization. (2021, August 25). More than 700 million people with untreated hypertension. Retrieved from World Health Organization Web site:  
<https://www.who.int/news/item/25-08-2021-more-than-700-million-people-with-untreated-hypertension>
- [9] World Health Organization. (2021, August 25). Hypertension. Retrieved from World Health Organization Web site:  
<https://www.who.int/news-room/fact-sheets/detail/hypertension>
- [10] 三軍總醫院。(2022年10月25號)。促進健康體能的方法。取自網址：  
<https://www.tsgh.ndmctsgh.edu.tw/unit/102337/22963>
- [11]成謝鋒，戴世成，趙鵬軍(2020)。基於心音信號的一種血壓評估方法。物理學報，69(14)。doi:10.7498/aps.69.20200254。取自網址  
<https://wulixb.iphy.ac.cn/article/doi/10.7498/aps.69.20200254>
- [12] 林新進(2009年7月)。淺談高血壓。高醫醫訊月刊，29(2)，11。取自網址  
<https://www.kmuh.org.tw/www/kmcj/data/9807/11.htm>
- [13] 高盈穎(2021年9月22號)。智慧手錶測血壓 精確度還要再等等。DIGITIMES。取自 DIGITIMES 網址  
[https://www.digitimes.com.tw/iot/article.asp?id=0000619443\\_0TW344NC5D30OB2LRXAO](https://www.digitimes.com.tw/iot/article.asp?id=0000619443_0TW344NC5D30OB2LRXAO)

H



- [14] 高點醫護網（2018）。內科一心音，臨床醫學教室。取自高點醫護網網址  
<https://doctor.get.com.tw/m/Journal/detail.aspx?no=406299>
- [15] 衛生福利部國民健康署（2020年10月15號）。世界高血壓日 依循 3C 原則  
守住 120/80 血壓的最佳防線。  
<https://www.hpa.gov.tw/Pages/Detail.aspx?nodeid=4141&pid=13232>
- [16] 衛生福利部國民健康署（2018年5月10號）。高血壓。取自衛生福利部網址  
<https://www.hpa.gov.tw/Pages/List.aspx?nodeid=1463>

## 【評語】 190002

此作品把心音轉成時序頻譜圖再利用 CNN 來訓練模型和預測血壓落在五個類別（區間）的哪一各類別。技術上，此作品是 CNN 的一種常見應用。商業上，要持續量一段心音才能轉換為血壓比起直接用血壓計來量血壓不見得有競爭力。另外，目前 AI 模型預測的效能是以對五個血壓類別來計算 Accuracy/F1 等分數而不是對真正血壓數值（ground truth）來計算數值誤差（如 RMSE），因此效能評估較不精準，建議同時使用 RMSE 來呈現和進行比較。