

臺北市 112 年度中等學校學生科學研究獎助

研究作品說明書封面

組 別：高中職組

科 別：應用科學科

作品名稱：利用咳嗽聲音自動辨識肺部疾病之系統開發

學生姓名：李博凱、賴東碩

作品編號：A53

製作說明：

1. 組別、科別、編號等須與複審通過公告之內容相符，作品名稱即研究計畫名稱。
2. 封面編排由參展作者自行設計。
3. 請與研究計畫作品說明書一起裝訂整齊，連同參加其他競賽獲獎申報書交由所屬學校統一送至北一女中設備組。

臺北市 112 年度中等學校學生科學研究獎助

研究作品說明書內容

作品名稱：利用咳嗽聲音自動辨識肺部疾病之軟體開發

摘要.....	4
壹、研究動機.....	4
貳、研究目的及研究問題.....	5
參、研究設計.....	6
肆、研究結果.....	28
伍、討論.....	34
陸、結論.....	36
柒、參考資料及其他.....	37

摘要

本研究主要的目的為當一筆咳嗽音來時，同時對多種肺部疾病進行辨識，除了偵測該產生咳嗽音的人是否患有肺部疾病外，還可以知道患有哪種疾病，藉以達成「診斷」效果，本研究將以隱藏式馬可夫模型(Hidden Markov Model,HMM)和梅爾倒頻係數(Mel-Frequency Cepstral Coefficients, MFCC)特徵提取來達成，並以超引數優化等方式優化結果

壹、研究動機

由於近年新冠病毒疫情肆虐，導致肺部相關疾病的意識受到提升，因為肺部相關疾病，可能和新冠病毒產生合併症（Comorbidity）或是併發症[1]，提升了重症的機率。以哮喘為例，哮喘的病因為過敏發炎而導致支氣管收縮，導致呼吸不易，它需要透過吸入皮質類固醇治療。雖然沒有吸入皮質類固醇治療的哮喘患者和一般患者的重症率相差無幾，但吸入皮質類固醇的成年患者有較一般患者高出 50%的機率轉為重症，而在吸入皮質類固醇的青少年則有比一般患者高出一倍的重症風險，而哮喘患者中有接受哮喘治療的比例較高，因此可以推知哮喘患者重症的機率比一般患者還高[2]。另一種類似的疾病是哮吼，同樣於哮喘是因肺部阻塞而引起，不過它是因為病毒而導致，所以在新冠病毒的肆虐下，它的發病機會被提[3][4]，也會導致新冠病毒的初始治療難度提升和轉重症的機率提升，而且最近常常有嬰兒因為哮吼和新冠病毒的共同發病而不幸去逝[10][11]。由以上結論可推知新冠肺炎可能誘發肺部相關疾病

而導致重症。而最近有研究指出一種新觀念「長新冠」，其指的是在罹患新冠肺炎後而產生的後遺症，此後遺症也會導致肺部相關疾病的發作機率提升。

貳、研究目的及研究問題

一、目的

由於咳嗽聲是一種生物標記[5]，透過觀查健康和病人波型的區別，可以透過它來偵測疾病，因此在國外已有研究使用咳嗽聲來對單一肺部疾病[6]和精神疾病[7]進行辨識，但普遍辨識率不高，或者是辨識率高但計算消耗量大，需要的測試資料也很多。而且僅對單一咳嗽疾病進行辨識無法達到「診斷」的功能，只能判定是否患有該種疾病，因此我們想要做出能夠一次對多種肺部疾病進行判定是否患有該疾病的模型使用較少的計算量，而且獲得較高的辨識率。

二、文獻回顧

(一)1996 年[8]該篇文章提到如何錄製咳嗽的短促聲以避免聲音中的特徵

不會被忽略，並透過觀察聲波圖形，觀察到一般健康的咳嗽聲音，是由兩個較大強度且不同的波形組成，第一個波形是在空氣道的窄縫和分叉處產生的，而第二個波形則是在喉部產生，但如果患有嚴重發炎那麼聲音的波形則會變大而且持續時間變長且波可能不僅由兩種不同的波型產生。因而發現咳嗽的波形有固定的變化模式，可以透過觀查波形的變化，得到局部病變嚴重程度。此外文章還提到第一個咳嗽聲音是由呼吸道的狹窄或分叉引起，第二個咳嗽聲音是由喉頭組織引起，中間的咳嗽

聲音是由痰液積聚在氣管引起，因此也可以觀查波型在哪個地方和健康的不一樣，得到肺部哪個位置出問題。

(二)2018 年[6]該篇論文使用梅爾倒頻譜係數(MFCC)和耳蝸圖像特徵

(CIF)分別對此聲音進行特徵提取，最後再以支援向量機(SVM)對哮喘和非哮喘的音訊進行分類。

參、研究設計

一、方法

先將音訊做前置處理，接著使用 MFCC 對特徵值進行擷取，再用 HMM model 對音訊進分析，獲得辨識率，最後再以不同參數輸入模型中，獲得多個模型及其辨識率，挑出辨識率最高時的 HMM 參數。

二、音訊前置處理

(一)測量語音端點：對不同的語音會有不同的門檻值(Threshold)，必須要到達一定的能量大小，才能將它視為一段有意義的音訊，否則它只是雜訊。

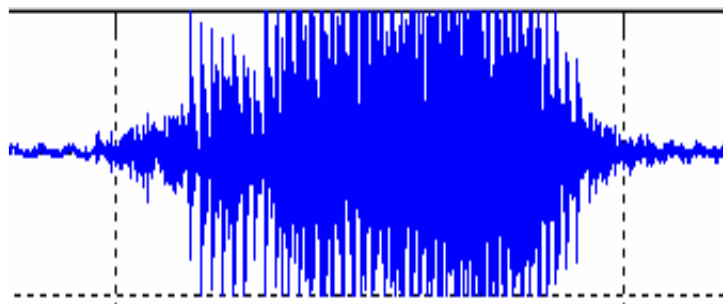


圖1: 語音端點偵測圖 (兩邊的虛線及為該訊號的兩端點)

(二) 擷取語音音框 :為了降低資料的運算量及信號的非穩態特性，我們會將音訊分成一個個小的音框，每個音框有256個點，而且每個音框互相重疊，目的是避免邊界信號的遺漏。

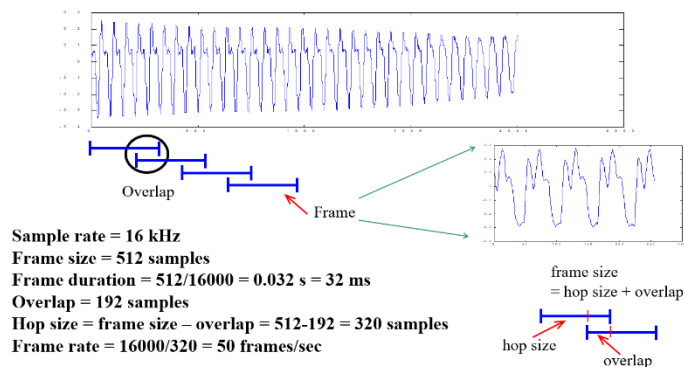


圖2: 擷取語音之音框圖 (圖的 Frame 為音框，overlap 為重疊部分)

(三)對高頻信號加強：人類發出高頻聲音時，會因發音過程，而降低了高頻聲音的強度，人耳會自動增益高頻聲音，但機器不會，所以為了彌補高頻聲強度，須對語音信號做預強調處理。

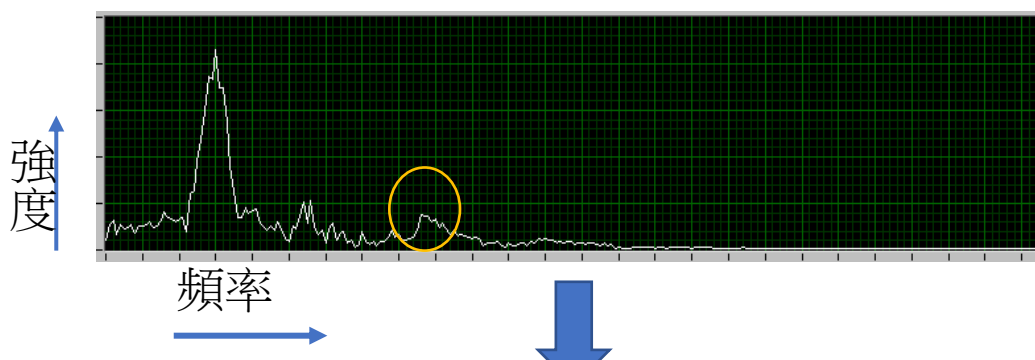
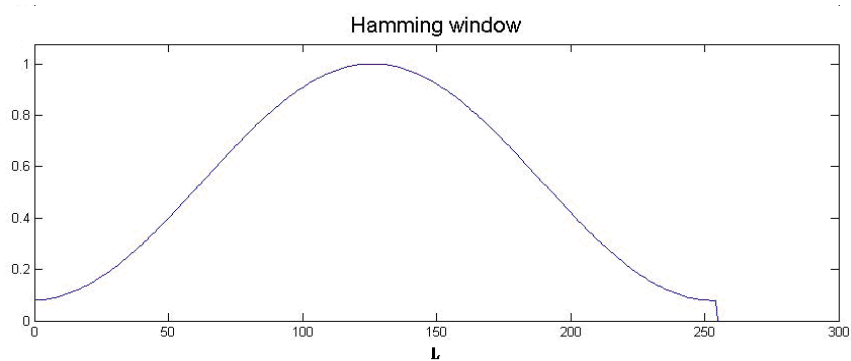


圖3：預強調前後頻譜變化

(四) 乘上漢明窗：為了避免相鄰音框邊界值的不連續，因為音框邊界值的不連續會讓本研究做傅立葉變換時多出一堆不存在的數值，此外還可以突顯音框中間數值的特性，因此本研究替每個音框乘上一個漢明窗。



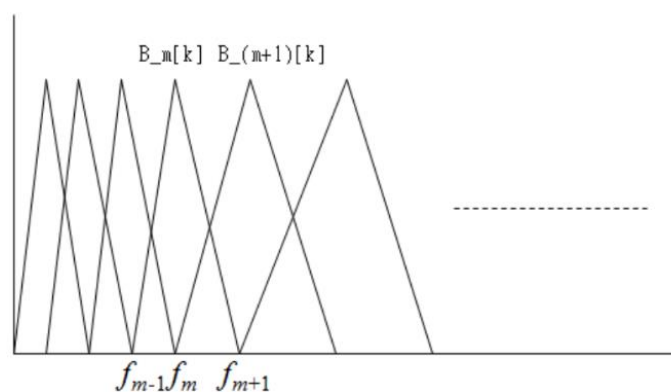
音框長度

圖4：漢明窗

三、梅爾頻率倒譜係數(MFCC)[9]

梅爾倒頻係數(MFCC)是針對人耳對聲音頻率的分辨能力所定義出來的，現在主流做語音辨識的都以梅爾倒頻係數來做音框的特徵值擷取，音框會先經過傅立葉轉換變成頻域後乘上梅爾頻率三角帶通濾波器累加之後取對數最後得到特徵值向量，其中特徵值向量的每個維度值是由一個三角濾波器做出，通常會擷取12-13維度的特徵向量。

(一)梅爾三角濾波器：此濾波器由26個左右的三角帶通濾波器組成的濾波器，每一個三角帶通濾波器偵測一個頻率範圍內的強度，每一個三角帶通濾波器的功能就如同偵測不同音調強度，假如有一個三角濾波器的端點在261.6Hz 左右該濾波器就是偵測Do 的強度。



圖五：梅爾三角濾波器(縱軸為強度、橫軸為頻率)

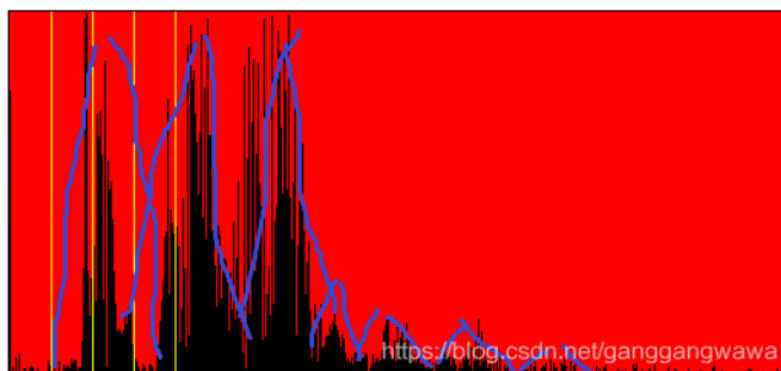
三角濾波器的數學表示方式：

$$B_m(k) = \begin{cases} 0, k < f(m-1) \\ \frac{2(k - f(m-1))}{(f(m+1) - f(m-1))(f(m) - f(m-1))} & , f(m-1) \leq k \leq f(m) \\ \frac{2(f(m-1) - k)}{(f(m+1) - f(m-1))(f(m) - f(m-1))} & , (f(m) \leq k \leq f(m+1)) \\ 0, k \geq f(m+1) \end{cases}$$

$B_m(k)$ 代表 m 號濾波器在 k 這個位置的頻率響應。

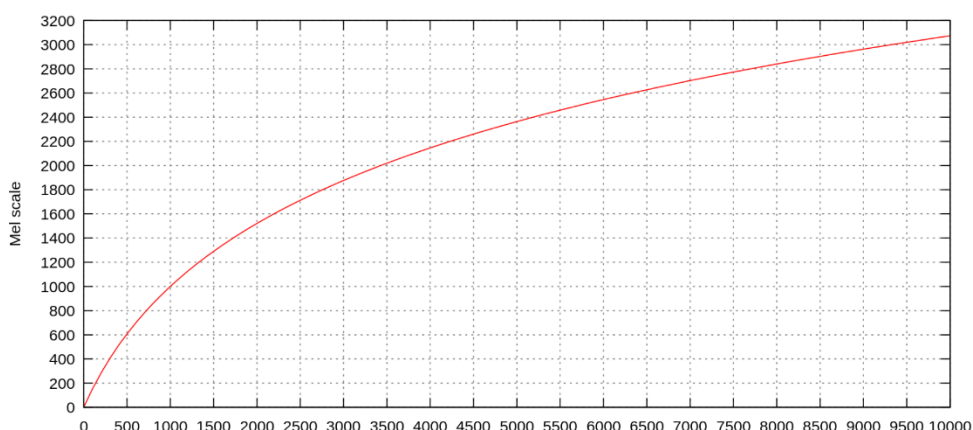
將能量譜通過一組梅爾尺度的三角形濾波器組，定義一個有 M 個濾波器的濾波器組（濾波器的個數和臨界帶的個數相近），採用三角濾波器，中心頻率為 $f(m)$ 。M 通常取22-26之間。每個 $f(m)$ 之間的時間會隨著 m 值的減小而變窄，隨著 m 值的增大而增寬。

傅立葉變換後的圖形：可以透過傅立葉變化後的圖形來解釋上文所述，可以觀察到傅立葉變換後的頻譜圖，是由一個個三角形組成，與梅爾三角濾波器形狀類似，因此梅爾三角濾波器可以很好的去偵測不同的音調能量大小。同時本研究會發現有些音調的能量過於小，在計算特徵值時會因為它的太小而被忽略，但是其所帶來的特徵值同樣重要，所以本研究會對其能量做對數處理，使能量大與能量小的能量差距不會那麼大。



圖六：音訊圖(縱軸為強度、橫軸為頻率)

(二)梅爾頻率：他可以想像呈代表一般人耳對於頻率的感受度，因為人耳的感受度與和實際頻率是呈對數關係，意思是如果音階 Do 的實際頻率範圍大小為50hz，則 Rei 的頻率範圍大小會比50hz 來的大，所以我們之前所提到的三角濾波器才會由小到大。如下圖橫軸為實際頻率，而縱軸的刻度區間就如同一個音階，例如0-200,200-400,400-600…，如同 Do Rei Mi…，縱軸的刻度被定義為梅爾頻率，梅爾頻率的出現可以幫本研究把原本呈對數關係的人耳感覺與實際頻率，改成為人耳感覺對梅爾頻率的線性關係表達。



圖七：梅爾頻率與真實頻率轉換圖

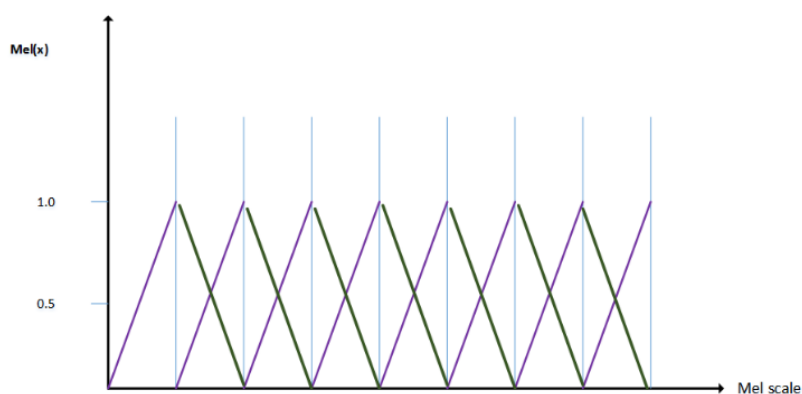
梅爾頻率數學式表達，其中 $mel(f)$ 是梅爾頻率 f 是原始頻率：

$$mel(f) = 2595 * \log_{10}(1 + f/700)$$

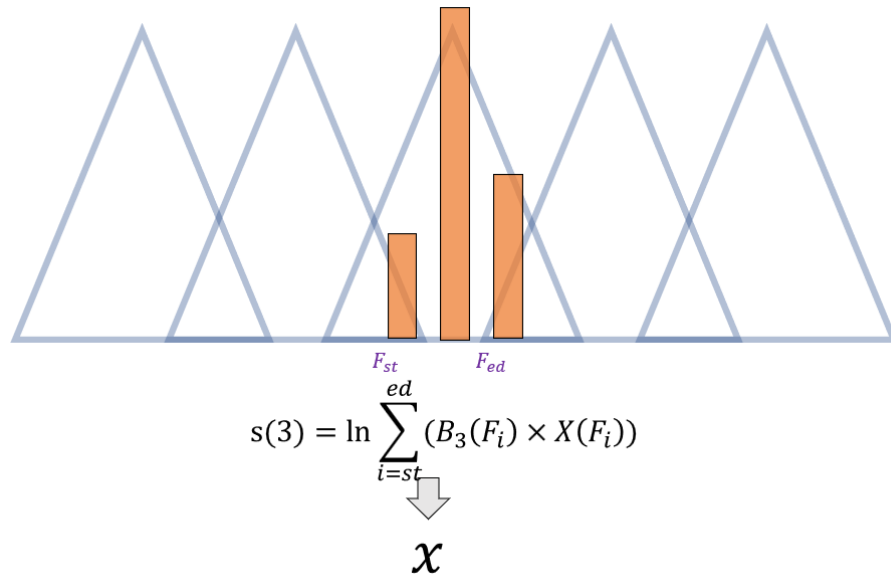
(三)梅爾三角濾波器的寬度會越來越長：如果本研究取梅爾頻率中的200-600的區間，並把這一段區間想像成一個音階，可以發現其所對應的實際頻率區間大約為0-500，

比區間大小同樣為400且梅爾頻率為2000-2400所對應的實際頻率區間3500-5000還要小的許多，所以梅爾三角濾波其才會呈現三角形大小由小到大的趨勢。

(四)梅爾頻率上的三角濾波器：透過梅爾頻率本研究就不需要製造大小區間不一樣的梅爾三角濾波器，只要做出等距的梅爾三角濾波器即可，因為梅爾頻率和人耳所聽到的感受度是呈線性的關係。因此使用梅爾頻率來製造梅爾三角濾波器可以方便程式實作以及避免浮點數誤差。



圖八：梅爾頻率上的三角濾波器



圖九：單個濾波器計算公式

此為以第三個濾波器所獲得的值，其中 $s(3)$ 代表三號濾波器輸出的能量， $B_3(F_i)$ 代表三號濾波器在 F_i 這個位置的頻率響應，也就是在 F_i 上的三號濾波器所對應的 Y 值， $X(F_i)$ 代表頻率 F_i 所具有的強度。

四、向量量化

在經由傅立葉變換，每一個音框的音訊會被轉為一 MFCC 特徵長向量，約12-13 維度。為了加快本研究的訓練速度和減少記憶體占用，本研究使用向量量化對其進行優化。向量量化是先將資料點進行分群，並把該資料點的長向量用該群心的編號取代，假設有某點被分第1群，那麼本研究會把向量改為一個值1表示，以此類推和第一群相似的資料點它特徵值都會變成一個值1。本研究將使用 K-means 演算法[14]進行分群。

五、隱藏式馬可夫模型(HMM)[13]

(一)使用隱藏式馬可夫模型的目的

我們使用隱藏式馬可夫模型有幾項好處，首先因為咳嗽聲在網路上資料量很少，此外本研究還要確保資料庫的可信度，所以此資料庫也必須曾被論文拿來做實驗，導致最後剩下的資料變得少之又少，但此模型不像深度學習的模型需要大量資料訓練資料才能獲得好結果，透過數學方法可以大幅度減少所需的訓練資料，減少訓練的時間。

(二) 隱藏式馬可夫模型(HMM)介紹

此模型的核心概念為在於透過可觀測有時間順序的現象，得到它的隱藏狀態。例如溫度是可觀測的「現象」，而春夏秋冬是勢不可觀測的「狀態」，如果溫度上升則代表有很高的機率代表夏季到來，而這就是使用 HMM 概念的一個例子。

(三) 將隱藏式馬可夫模型應用到本研究的方法[12]

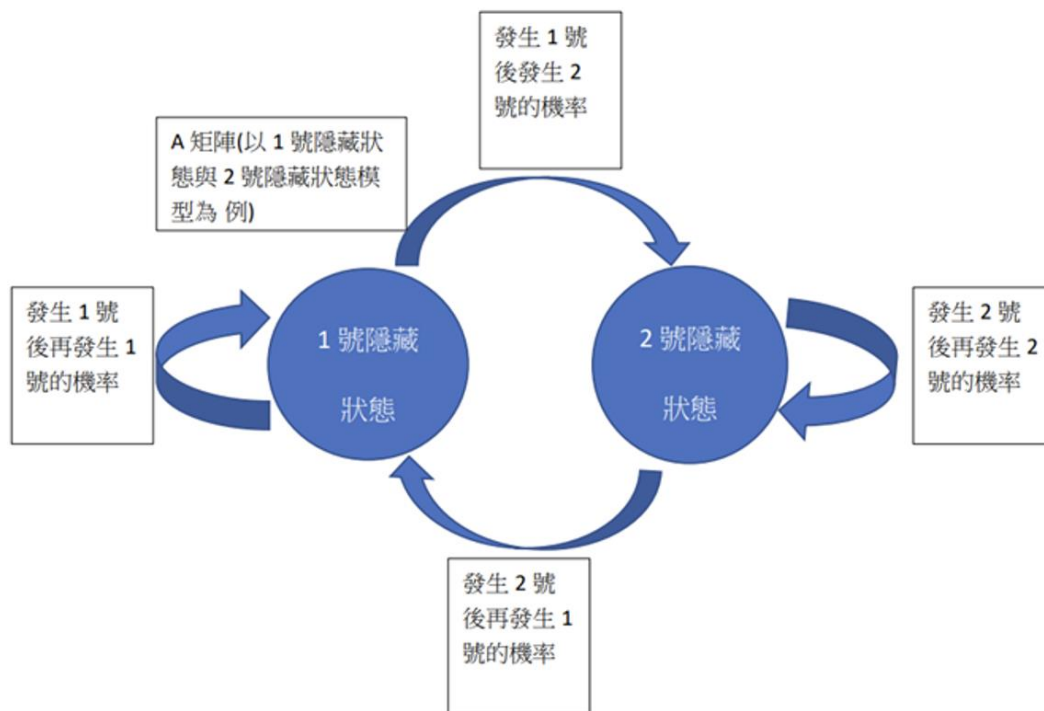
本研究與普通隱藏式馬可夫模型的應用不太相同，如果以上方例子再做舉例我們是將所有為夏天的的資料丟入一個為偵測與夏天相似度的模型，並鬥過這些資料將其訓練出來，其中相似度會以機率的型態表示，當有新的資料丟入後，本研究會透過隱藏式馬可夫模型的算法計算，獲得新的資料的可觀測狀態和新的資料的隱藏

狀態，把這兩個資料丟入偵測與夏天相似度的模型做計算，而得到相似度，所以本研究所使用的隱藏狀態不具有實質的物理意義，純粹僅是計算機率，並記錄觀測值和隱藏狀態的關係。

(四) 隱藏式馬可夫模型基礎架構

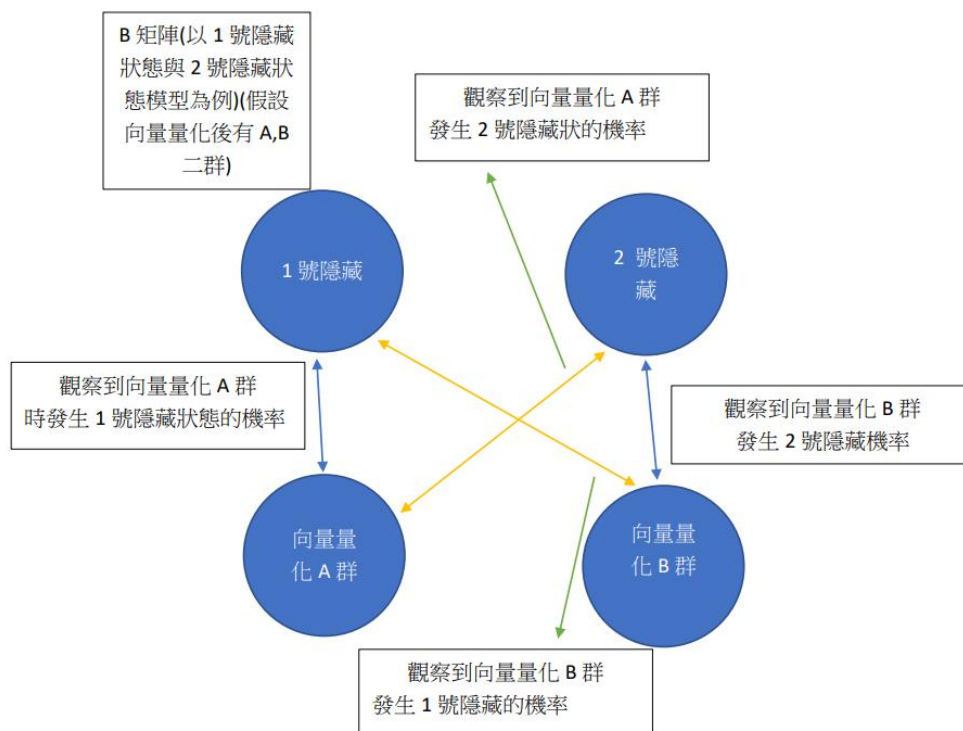
隱藏式馬可夫模型由三個矩陣組成分別為 A 矩陣 (狀態轉移機率矩陣)、B 矩陣 (條件概率矩陣)、 π 矩陣(初始機率矩陣)，A 矩陣所代表的意義為隱藏狀態之間的狀態轉移機率，B 矩陣所代表的意義為在已觀測到某種現象下，得到的隱藏狀態的機率， π 矩陣為初始並未觀測到任何結果前的機率。我們用將以圖解釋本研究 A、B 矩陣的意義

(五)A 矩陣



圖十：A 矩陣式意圖

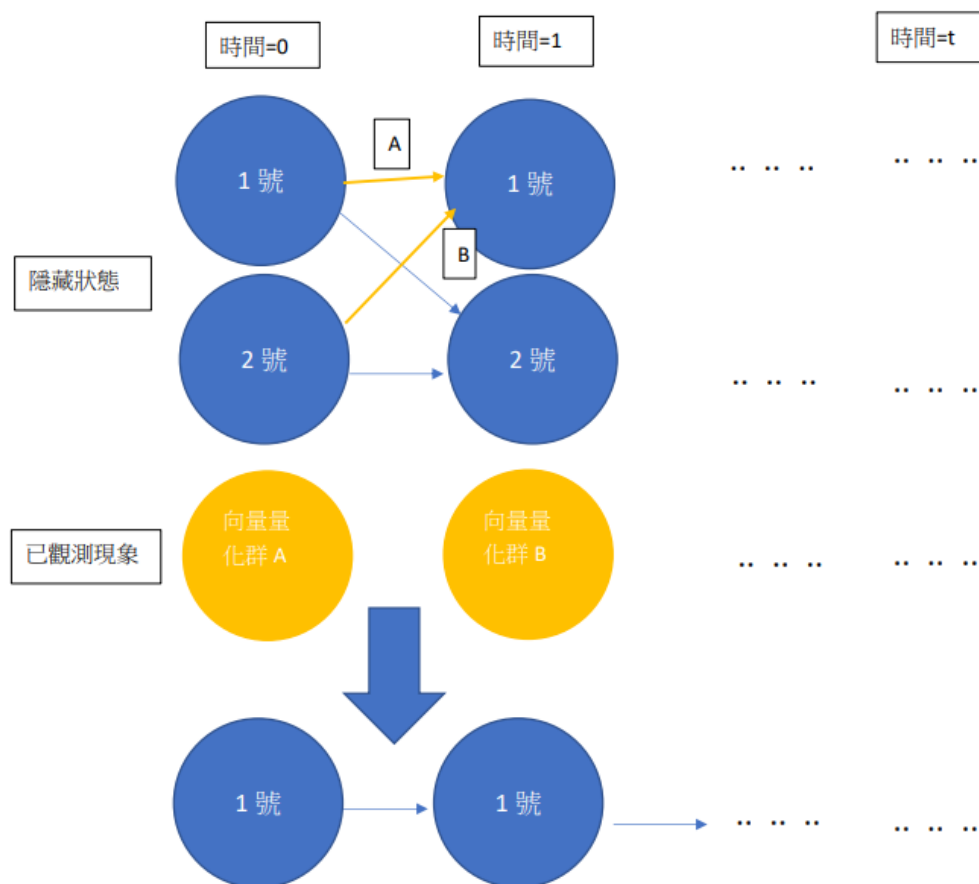
(六)B 矩陣



圖十一：B 矩陣式意圖

(七)獲得最大機率路徑：接下來需要透過已有的機率矩陣得出一條最

大路徑的機率，我們將以做圖和下方文字解釋。



圖十二：最優機率路徑估測

這張圖中，上方藍色圈圈為隱藏狀態，下方的黃色圈圈為向量量化群後該資料點所分配到的群心(可觀測現象)。

先關注黃色兩條線 A 和 B，可以得知發生路經 A 的機為：

$$P(\text{初始為1號機率})(\pi \text{ 矩陣之值}) \times P(\text{向 量 量 化 群 B} | \text{1號})(B \text{ 矩陣之值}) \times P(\text{...})$$

前一個為1號機率 | 下一個為1號機率)(A 矩陣之值)

而 B 路徑的機率為：

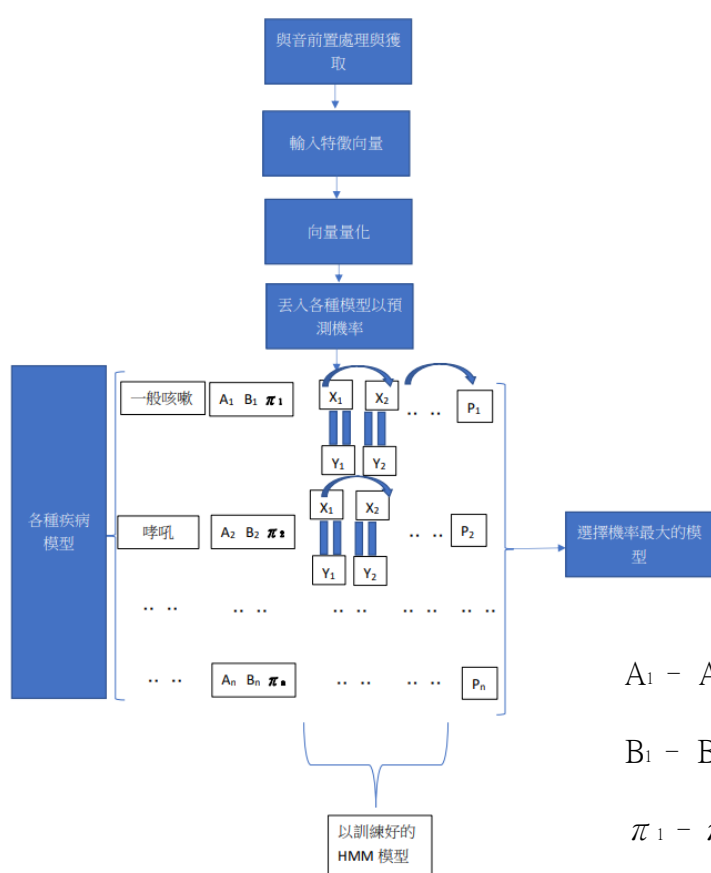
$P(\text{初始為2號機率})(\pi \text{ 矩陣之值}) \times P(\text{向 量 量 化 群 B | 2號})(B \text{ 矩陣之值}) \times P(\text{前一個為1號 | 下一個為2號})(A \text{ 矩陣之值})$

說明：如果 $P(A) > P(B)$ 那麼隱藏狀態1, 1, …這條隱藏狀態序列會比1, 2, …這條序列的機率還好，因此我們會選擇機率較大的那條作為我們的結果，並以此更新矩陣中的機率。

(八)訓練流程：整體的訓練流程為先給予不同疾病不同的偵測模型，然後在一個疾病模型中隨機的初始化每個矩陣的機率，接著取一疾病音訊的一個音框做傅立葉變換和 MFCC 後，注意一開始的所有矩陣皆為隨機值，‘接著透過這些矩陣獲得最大機率路徑，接著透過此路徑修改矩陣中的機率。修改方式如下，我們假設要修改 A 矩陣中1號隱藏狀態轉移到2隱藏狀態的數值我們需要統計所有1號隱藏狀態轉移的情形，包含1號隱藏狀態轉移至1、2、3、4…隱藏狀態的總數，透過總數分之1號隱藏狀態轉移到2隱藏狀態的數量得到機率，接著再存入矩陣中，B 矩陣的更新亦然。然後就把該音訊的所有音框經由此方法訓練，再把同樣疾病的音訊丟入以訓練該疾病所對應

的模型，其他疾病模型也以相同的方式訓練。

(九) 模型預測音訊所屬疾病: 當一筆新的音訊丟入後我們會丟入各個模型中得到最大機率路徑，本研究會選擇所有模型中機率機率路徑最大的那一條來找出其所對應的疾病來當作我們的預測值。如果以圖表示則為下圖



$A_1 - A_n$ 代表偵測不同的疾病的狀態轉移機率

$B_1 - B_n$ 代表偵測不同的疾病的狀態輸出機率

$\pi_1 - \pi_n$ 代表初始狀態機率向量

$X_1 - X_n$ 代表自訂隱藏狀態

$Y_1 - Y_n$ 代表量化結果

$P_1 - P_n$ 發生該疾病的機率

圖十三：模型預測音訊所屬疾病

六、超引數優化與 tenfoldset

(一) 超引數優化：當我們將模型訓練完後可能對此結果不甚滿意，所以我們會對此模型調整他的參數來達成更好的效果，本研究將調整上述的群心數量，和隱藏狀態數

量，可以調整隱藏狀態數量是因為隱藏狀態在本研究並不具有實質物理意義，接下來我們會把這些具有不同參數的模型稱作「實驗點」。

(二) tenfoldset 技術：是在機器學習中常用的一種交叉驗證技術，它將完整數據集分為十個大小相似的互不重疊的數據集，接著再將其中九個來做為訓練集，剩下一個做為測試集，再將此方法重複十次，每次測試集都選與之前測試集不同的資料集作為測試集，其餘則做為訓練集，透過此方法可以更好地評估模型的穩定性和泛化能力。

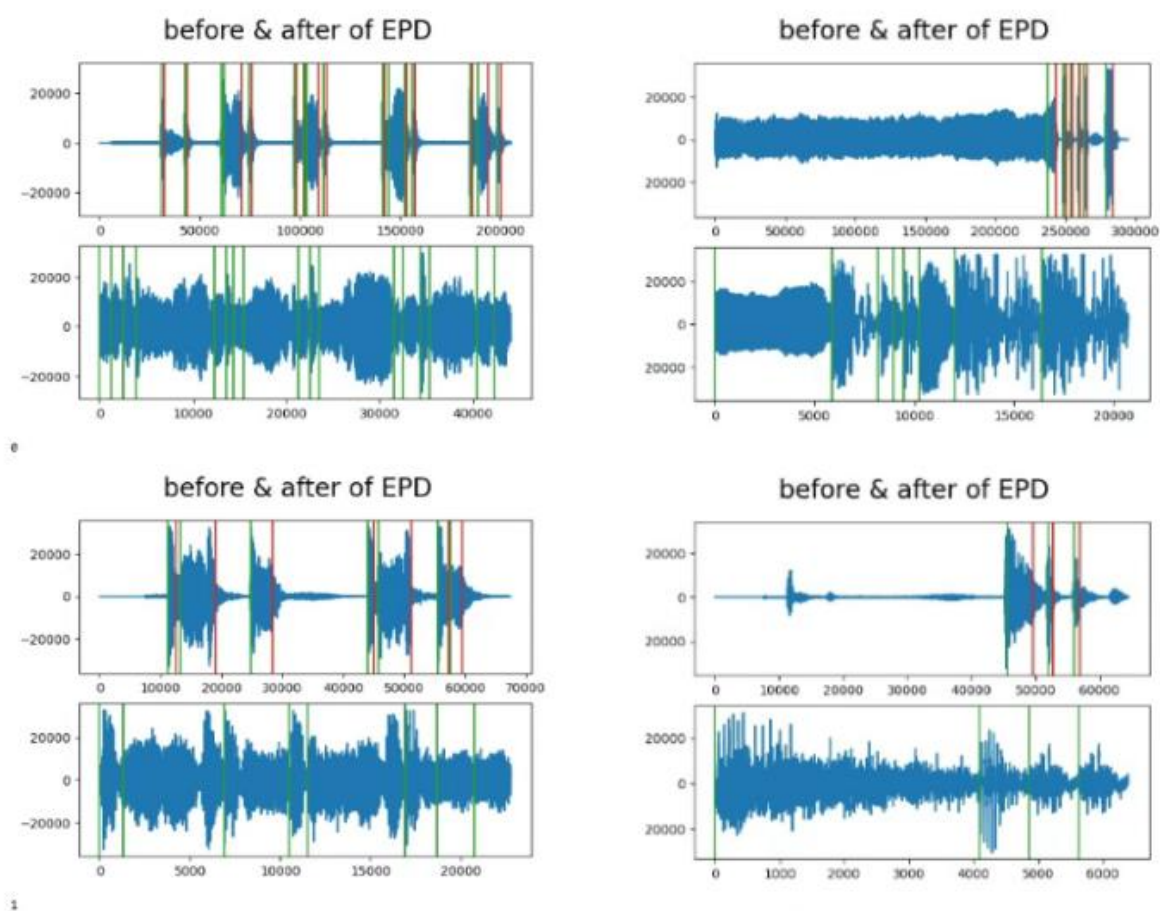
(三)資料集分割訓練流程： 本研究先將完整數據分割成7:3的訓練集和測試集，並透過這些資料進行超引數優化實驗，接著我們會選前五個具有較高辨識率的實驗點，一一挑出每個所具有的參數，接著我們會透過 tenfoldset 訓練出十個不同的新模型，做十次實驗，每次實驗會將其中一個數據集當作測試集，其他的則當作訓練集，再透過測試集獲得辨識率，這樣的情形重複做十次，最終將得到的所有辨識率取平均，再比較我們所取出的實驗點何者辨識率最大，並將其選出。此外雖然本研究的資料集並不平均，但本研究在程式實作時，是將相同疾病的資料訓練偵測該模型的疾病，所以不會有資料偏斜的問題。

七、波形圖與端點測量後的波形圖

上述有提及端點測量為去除聲音無信號的技巧，可以把沒聲音的部分去除也從而加

快我們的訓練速度，不過在本研究使用此技巧可能會有風險，因為兩個前後咳嗽聲音的間格可能可以做為一個特徵，如果使用此技巧則會使此特徵不見，所以本研究會以是否有經過端點測量作為實驗中的變量，來測試是會達到較好的效果。

下方分別為四種不同疾病的波型圖，左上方為哮喘，左下角是健康，右上角是 Covid-19，右下角是心臟衰竭，每個圖片的上方是尚未做端點測量(EPD)，下方為已經做完端點測量。



圖十四：音訊與端點偵測後音訊圖(縱軸為強度、橫軸為時間)

本研究所使用的資料來自 Elazığ 的 Firat 大學醫院肺科和心臟病學診所的842位被診斷患有 Covid-19、心衰竭或急性哮喘的病人，或被診斷為健康的病人，他們使用手機麥克風進行錄音資料並以 ogg、mpeg、mp4或 m4a 格式以44.1或48 kHz 的採樣速率儲存，總共蒐集了842個樣本，此資料集目前已經公開提供研究人員和開發人員使用，可以通過超鏈接訪問，此為資料集的超連結[16]。下圖為資料個數的描述：

疾病	疾病的資料數量
Covid-19	241
健康	247
心衰竭	244
哮喘	110
總共	842

圖十五：各項疾病資料數量資訊圖

訓練集與測試集的分割：本研究將訓練資料分割為，訓練資料有602測試資料有240，比例約為7:3，這樣的比例分配方式可以有效去除辨識率不穩定的部分。

九、模型優化與實驗設計

(一)優化方法：

1、是否經過端點測量

由於本研究使用端點測量技巧可能會有風險，因為兩個前後咳嗽聲音的間格可能可以做為一個特徵，如果使用此技巧則會使此特徵不見，所以本研究會實驗是否不做端點測量的效果會比作端點測量的效果好。

2、參數調整以優化結果

不同的群心數量和隱藏狀態數量會為有不同的辨識率，我們會調整這些參數來觀察這些參數對應辨識率的趨勢，本實驗的參數，包含隱藏狀態數量和群心的數量皆會設定成範圍為1-100之間的整數值，並且會以10個10個為一間隔的形式測試，即以1,11,21,31...的方式，因為實驗裝置運行速度不足以支撐每一個1-100的整數的組合進行測試。

3、MFCC 微分

針對 MFCC 進行微分（亦稱為一階差分）是一種提取語音信號動態特徵的方法。這個過程能捕捉語音信號隨時間變化的信息，從而提升如語音識別、發音分類和情感識別等任務的性能。

對 MFCC 進行微分的方法是計算相鄰幀之間的差值。通常，還會計算二階差分（即對一階差分再次進行微分），以捕捉語音信號的加速度信息。一階差分有時被稱為 MFCC 的"速度"特徵，而二階差分被稱為"加速度"特徵。

將原始的 MFCC 特徵與其一階和二階差分特徵結合在一起，可以形成一個更具表現力的特徵向量，從而提高語音處理任務的性能。

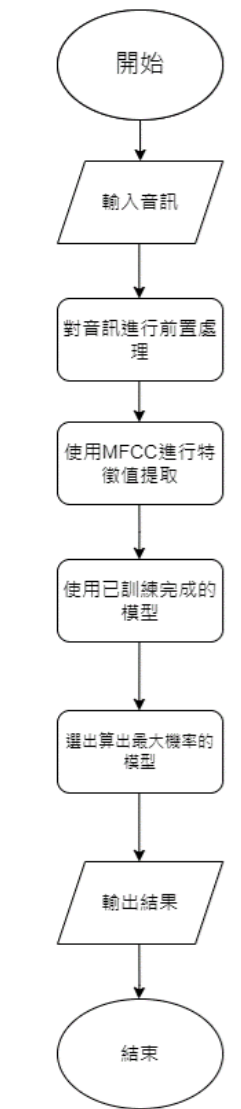
(二)實驗設計：本實驗將繪製兩張有無進行端點測量的立體圖，圖中畫出來會像一個平面的形式，該平面是由多個實驗點與結果連線後繪製而成，每個點有其對應的辨識率、隱藏狀態、群心數量，我們目標是找出前五個辨識率最高時所對應的隱藏狀態、群心數量，接著進行 tenfoldset 的交叉驗證，挑出五個點中平均辨識率最佳的，最後比較有無端點測量的差異。

(三)實驗描述

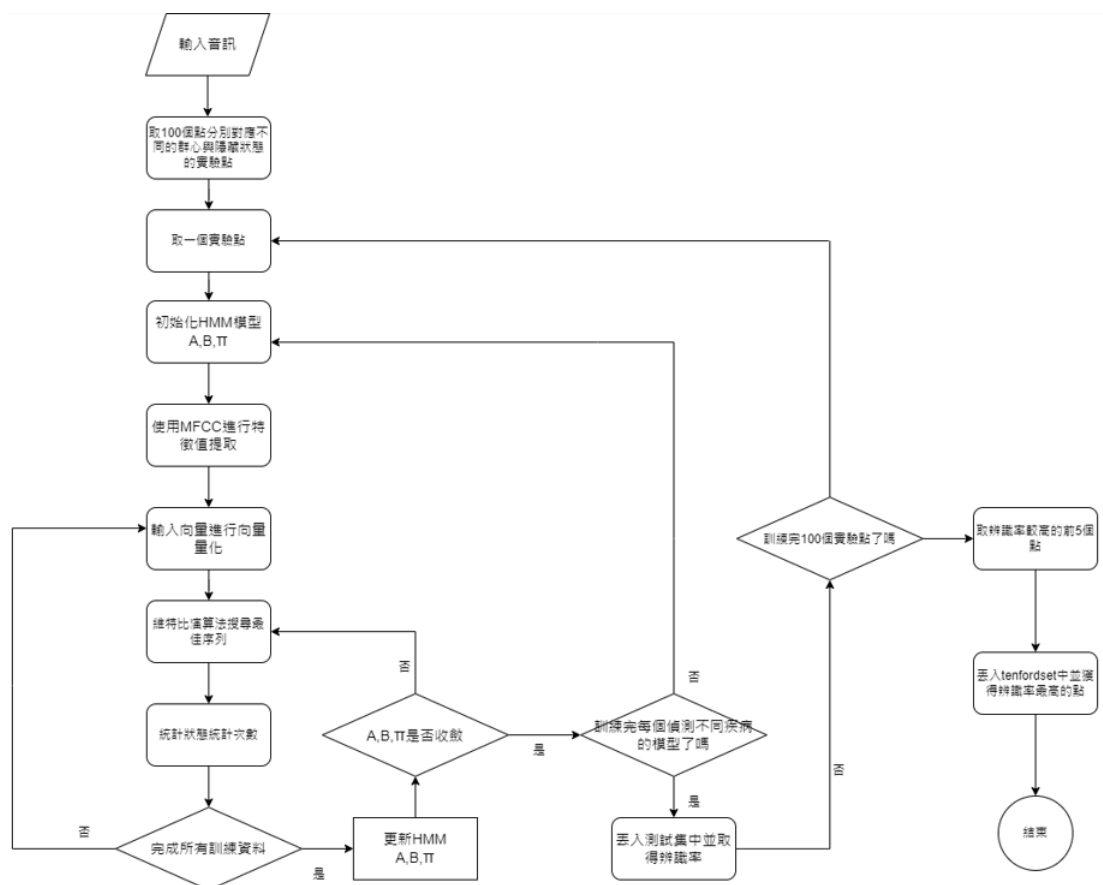
1. 實驗一：對有進行端點測量的音訊，經音訊處理後丟入馬可夫模型中進行訓練，並觀察其超引數優化的效果。
2. 實驗二：對沒有進行端點測量的音訊，進行同實驗一的處理。
3. 實驗三：比有無 MFCC 微分對模型訓練的影響。本實驗將使用未進行端點測量的音訊，分別對原始 MFCC 特徵和進行了一階和二階差分的特徵進行訓練，並觀察兩者在辨識率上的差異。

十、流程圖

(一)整體流程圖



(二)模型訓練流程圖



十一、評估指標

(一)效能評估：對於每個模型，我們將計算其在測試集上的未加權平均召回率(UAR)[19]，作為衡量其性能的指標。因為本研究使用的數據集並不平衡，例哮喘的疾病資料就很少，如果使用平均辨識率作為評判標準，則會無法獲得對較少資料的疾病的測試表現。使用召回率(recall)[20]的目的是為了評估模型從有患病的人中，能夠偵測到底多少人有疾病的百分比。

以下為計算公式：

$$UAR = \frac{1}{N} \sum_i^N recall(i)$$

i 為該疾病的種類

N 為疾病種類數量

$recall(i)$ 為對 i 這個疾病的召回率

(二)訓練時間：對於每個模型，我們將記錄其訓練時間，並比較不同模型的訓練效率。

(三)資源需求：對於每個模型，我們將記錄其所需的計算資源，包括內存和計算能力，

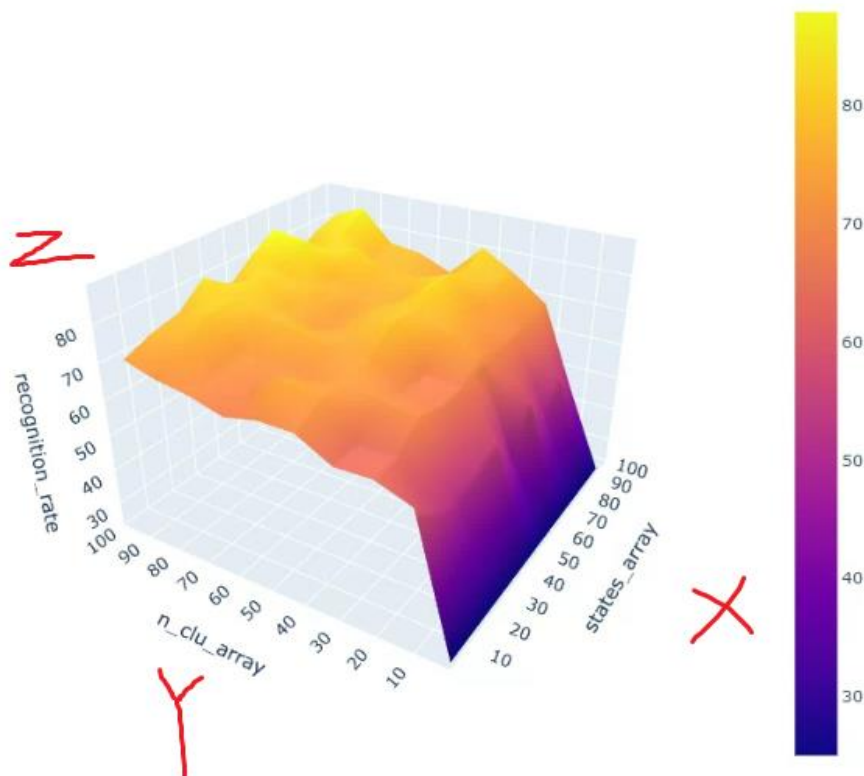
以評估其在不同硬體環境下的可用性。

肆、研究結果

(一) 有進行端點點測量的實驗：一般的語音辨識等實驗都會對音訊進行端點測量以刪除不含能量的音訊。

1、實驗數據圖

此立體圖為本專題超引數優化實驗結果 X 軸為隱藏狀態數量，Y 軸為群心數量，Z 軸為辨識率



圖十五：有進行端點點測量立體圖

前五個辨識率最高的實驗點分別為

	群心	隱藏狀態
實驗點 1	91	61
實驗點 2	81	91
實驗點 3	81	61
實驗點 4	81	81
實驗點 5	91	31

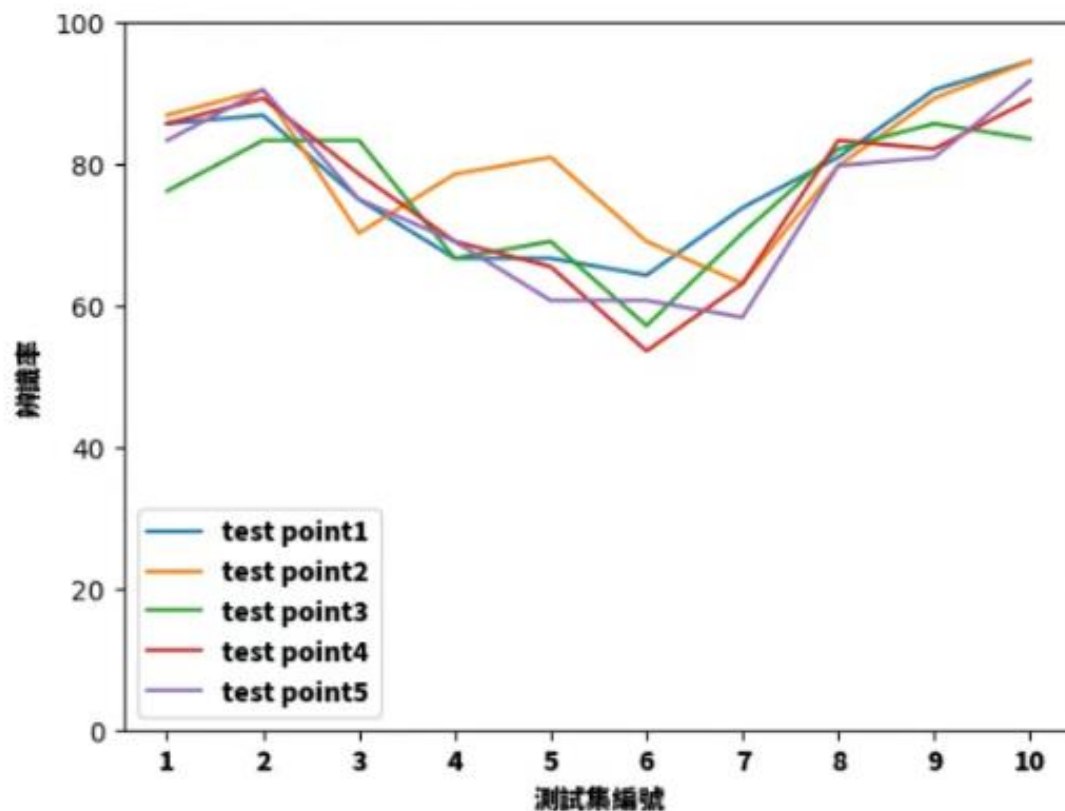
圖十六：前五個辨識率最高的實驗點資訊圖

2、前五個辨識率最高的實驗點進行 tenfoldset 結果

我會對每一個實驗點分別做以下實驗，先將完整的資料庫分為十個大小相等的數據集，每次實驗會將其中一個數據集當作測試集，其他的則當作訓練集，再透過測試集獲得辨識率，這樣的情形重複做十次，最終將得到的所有辨識率取平均

1.tenfoldset 辨識率圖

此圖表示不同的實驗點在以 x 軸上的編號為測試集時所表示的辨識率，每一條不同顏色的線為不同的實驗點，y 軸是辨識率，x 軸是此時辨識集的編號，例如以 5 號數據集為測試集，其他數據集為訓練集時，實驗點 2 號的辨識率為 80.95%



圖十七：有進行端點點測量 tenfoldset 辨識率圖

2 有端點測量.tenfoldset 平均辨識率

本實驗會以平均辨識率為標準，此指標象徵辨識率泛化能力的好壞，代表

當有一筆新的資料點進來時實驗點二會得到最好的結果。

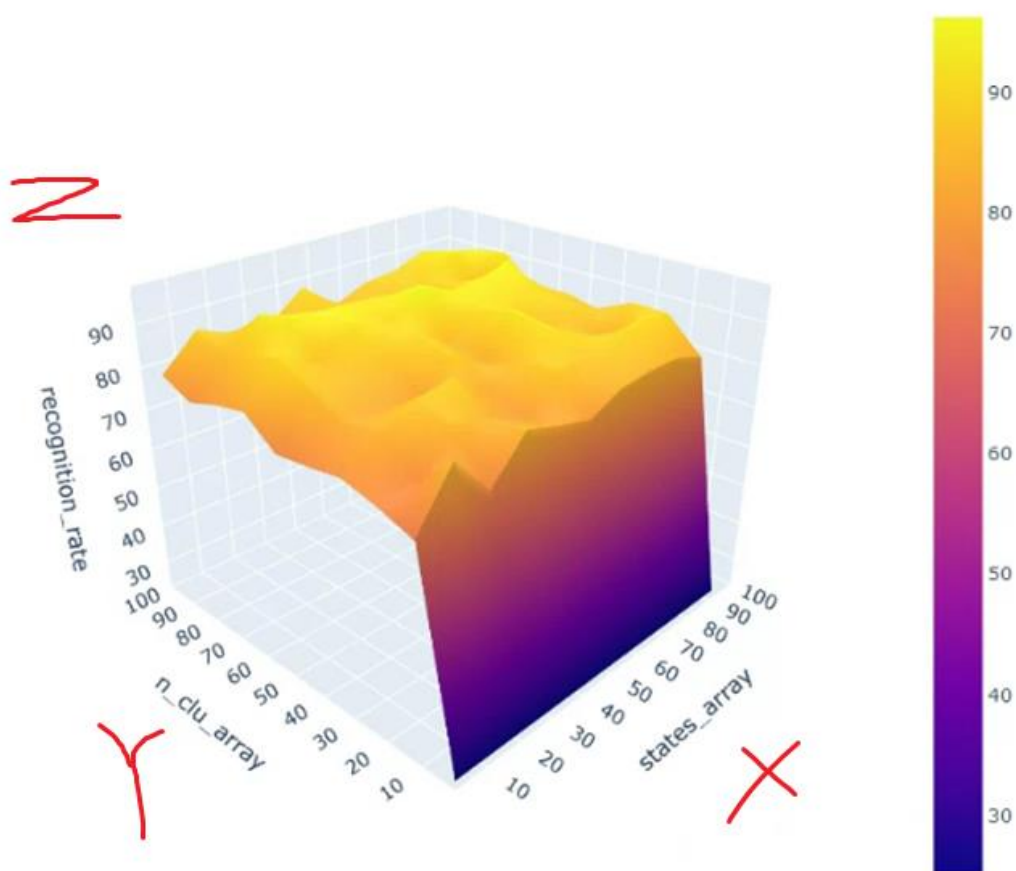
	平均辨識率
實驗點 1	78.50
實驗點 2	80.29
實驗點 3	75.74
實驗點 4	75.9
實驗點 5	75.01

圖十八：有進行端點點測量 tenfoldset 平均辨識率圖

(二) 沒有做端點點測量的實驗：由於聲音與聲音之間有個，先對音訊進行端點測量的處理，

1、實驗數據圖

此立體圖為本專題超引數優化實驗結果 X 軸為隱藏狀態數量，Y 軸為群心數量，Z 軸為辨識率



圖十九：沒進行端點點測量立體圖

前五個辨識率最高的實驗點分別為

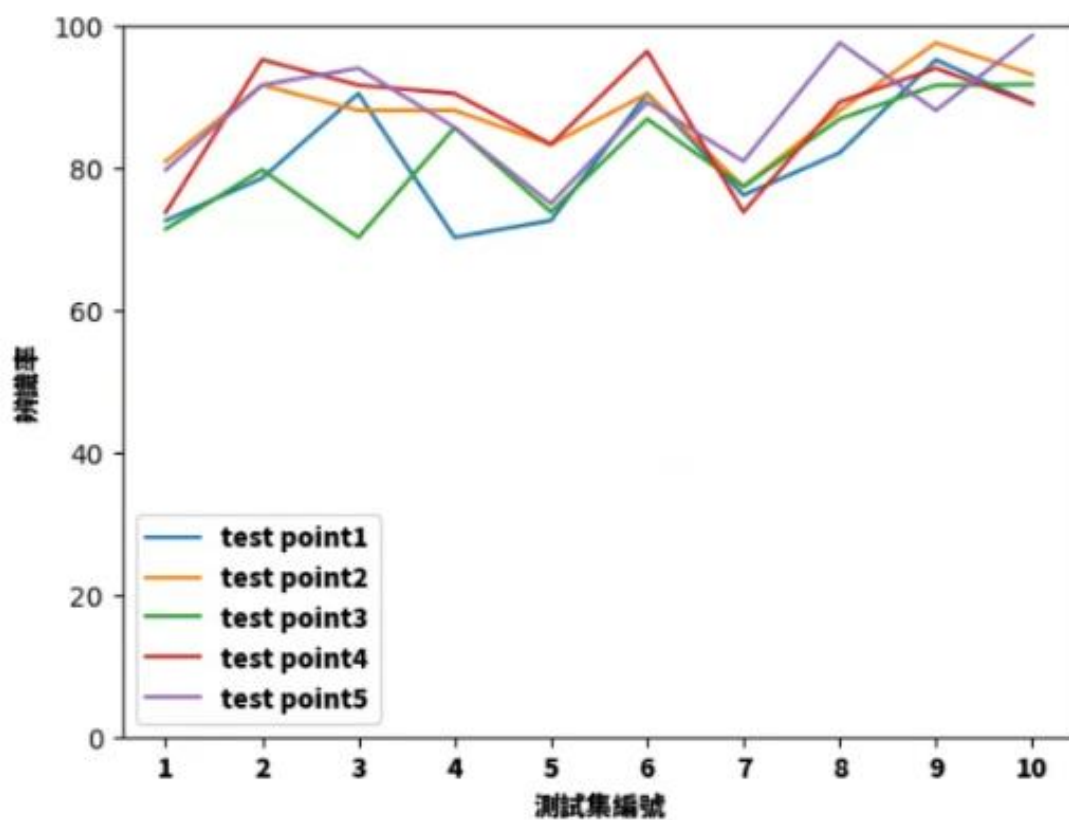
	群心	隱藏狀態
--	----	------

實驗點 1	61	51
實驗點 2	61	81
實驗點 3	71	21
實驗點 4	81	81
實驗點 5	71	91

圖二十：前五個辨識率最高的實驗點資訊圖

2、前五個辨識率最高的實驗點進行 tenfoldset 結果

1.tenfoldset 辨識率圖



圖二十一：沒進行端點點測量 tenfoldset 辨識率圖

2.無端點測量 tenfoldset 平均辨識率

	平均辨識率
實驗點 1	81.76
實驗點 2	87.89
實驗點 3	81.56
實驗點 4	87.71
實驗點 5	88.08

圖二十二 :沒進行端點點測量 tenfoldset 平均辨識率圖

(三)比較有無 MFCC 微分對模型訓練的影響

1、無使用 MFCC 微分之數據：

模型編號	群心	隱藏狀態	端點測量	UAR(10-Fold)
1	91	61	有	83.33
2	81	91		82.92
3	81	61		76.67
4	81	81		76.25
5	91	31		83.75
6	61	51	無	83.75
7	61	81		91.25
8	71	21		84.58
9	81	81		88.33
10	71	91		88.33

圖二十三：、無使用 MFCC 微分之數據

2、有使用 MFCC 微分之數據：

模型編號	群心	隱藏狀態	端點測量	UAR(10-Fold)
1	91	61	有	68.33
2	81	91		57.55
3	81	61		50.83
4	81	81		58.33
5	91	31		57.55
6	61	51	無	86.25
7	61	81		91.25
8	71	21		90.00
9	81	81		85.42
10	71	91		88.75

圖二十四：、使用 MFCC 微分之數據

伍、討論

一、比較有無做端點測量的效果差異

一般情況下的語音相關實驗都會做端點測量，可是在本專題下使用端點測量則會導致語音與語音之間的停頓特徵消失，導致辨識率比沒做端點測量時平均來說都少了 10% 左右。

二、超引數優化討論

發現立體圖辨識率在參數數值小時辨識率大小會隨著群心大小上升，而當參數數值到一定大小時辨識率則會在特定的參數組合上表現較大或較小的辨識率，在訓練集與測試集比例為 **7:3** 且沒做端點測量時，優化後的結果最高可達 95.42%。

三、tenfoldset 實驗結果討論

此實驗結果幫助評估此模型遇見從未看過資料時表現的好壞，例如在有做端點測量時，黃色實驗點在訓練集與測試集比例為 **7:3** 上的表現並非最優，但是在遇見從未見過的資料時，整體的平均表現優於其他實驗點。

四、MFCC 微分實驗結果討論

使用 MFCC 微分之優化方法可以獲取較多訊息，所以使 UAR(10-fold)比沒有使用的多約 1-2%。

五、CNN 與本研究方法之比較圖

分別對以下三分文獻與本研究做比較：

研究一：DETECTING COVID-19 FROM BREATHING AND COUGHING SOUNDS USING DEEP NEURAL NETWORKS 2020

研究二：The INTERSPEECH 2021 Computational Paralinguistics Challenge: COVID-19

Cough, COVID-19 Speech, Escalation & Primitives 2021

研究三：Roneel V Sharan, Udantha R Abeyratne, Vinayak R Swarnkar, Paul Porter Automatic Croup Diagnosis Using Cough Sound Recognition(2018)Automatic Croup Diagnosis Using Cough Sound Recognition

	uar(hold-out)	準確率(out fold)	uar(10-fold)	單一或多類辨別	方法	電腦成本	資料集收集處	能辨識的疾病類別
研究一	74.9	73.1		多	DNN	多	University of Cambridge	COVID-19、哮喘、健康
研究二	87.5	69.3		單一	RNN	多	University of Cambridge	COVID-19
研究三	95.01	92.31		單一	SVM	少	A hospitals in Australia	哮喘
本研究	91.25	91	91.25	多	HMM	少	Firat University Hospital, Department of Pulmonology and Cardiology, Elazığ, Turkey	COVID-19, 哮喘、健康、心衰竭

圖二十五：CNN 與本研究方法之比較圖

陸、結論

一、現有的研究多著重於單一的疾病辨識，雖然辨識率高但無法獲得廣泛的應用，也有的是直接使用深度學習達成多功能的應用，不過此方法需要大量的資料與也要透入相當大的計算成本，本專題的方法可以解決上述的問題，並有相當好的辨識率。

二、本專題也使用不做端點測量以及超引數兩種優化方法來提高辨識率，和用 tenfoldset 來加強本專題在未來應用的可信度。

三、未來展望與未來應用

(一)在特徵提取部分本專題希望可以做到特徵優化，因為本專題的特徵是隨機取 12-13 維度，而實際的特徵值多達 20-30 維度，希望透過此方法優化辨識率

(二)將本專題研究做成網站線上形式，讓有需要的人能上該網站自己檢測是否患有該疾病。

(三)快篩時需要拭子深入鼻子內部，常會讓人感到不舒服，可以將本專題的程式拿來測試，如果發現是陽性，再做快篩以確認。

(四)在做全面普篩時由於健康的比率佔大多數，而真正有疾病比率較小，如果未來要應用在普篩上，可以使用 EER Rate[15]來對模型建立門檻值做辨識。

柒、參考資料及其他

[1] (2015) 共病症 與 併發症 (Comorbidity & Complication)

[2] Ted Dolby, Vahe Nafilyan, Ann Morgan, Constantinos Kallis, Aziz Sheikh,(2022)

《Thorax》 Relationship between asthma and severe COVID-19

[3] Sujit Sharma 1, 2, 3 , Beesan Agha 1, 2 , Carlos Delgado 1, 2 , Karen Walson 4 , Charles Woods 5 , Mark D Gonzalez 6 , Robert Jerris 6 , Gregory Sysyn 4 , James Beiter 1, 2 , Satoshi Kamidani 7, 8 , Christina A Rostad 7, 8(2022) Croup Associated With SARS-CoV-2: Pediatric Laryngotracheitis During the Omicron Surge

[4] Taylor Brinton , Gregory Howell , Hashaam Arshad , Ain Ejaz , Ibrahim

Alakhras(2020) CLINICAL OUTCOMES OF PATIENTS WITH ASTHMA WITH COVID-19 PNEUMONIA

[5] Duffy, M. J., & Sturgeon, C. (2010). Biomarkers for early detection of cancer. *Seminars in cancer biology*, 20(1), 22-31

[6] Roneel V Sharan, Udantha R Abeyratne, Vinayak R Swarnkar, Paul Porter ()Automatic Croup Diagnosis Using Cough Sound Recognition(2018)Automatic Croup Diagnosis Using Cough Sound Recognition

[7] Pedro Gómez-Vilda, Victoria Rodellar-Biarge, Víctor Nieto-Lluis, Cristina Muñoz-Mulas, Luis Miguel Mazaira-Fernández, Rafael Martínez-Olalla, Agustín Álvarez-Marquina, Carlos Ramírez-Calvo & Mario Fernández-Fernández (2013) Characterizing Neurological Disease from Voice Quality Biomechanical Analysis

- [8] J Korpás 1, J Sadlonová, M Vrabec (1996) Analysis of the Cough Sound: an Overview
- [9] Bozkurt, B., & Arslan, L. M. (2015). Comparison of Mel-frequency cepstral coefficient feature sets for speaker identification. Turkish Journal of Electrical Engineering & Computer Sciences, 23(3), 854-863
- [10] 丁彥伶(2022) 親子防疫／兒童病例激增 幼兒呼吸、咳嗽發出「狗吠」聲最危險
- [11] 李沁原 李媛媛 田小银 张光莉 杨巧 刘玉琳 罗征秀(2020) 新型冠狀病毒肺炎對兒童哮喘控制水平及影響因素分析
- [12] Forney, G. D. (1973). The Viterbi algorithm. Proceedings of the IEEE, 61(3), 268-278
- [13] Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2), 257-286
- [14] Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. Journal of the Royal Statistical Society: Series C (Applied Statistics), 28(1), 100-108
- [15] Equal Error Rate (EER) <https://www.innovatrics.com/glossary/equal-error-rate-eer/>
- [16] 資料集超連結 http://web.firat.edu.tr/turkertuncer/acute_asthma_cough.rar
- [17] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).
- [18] Kaiming He, Xiangyu, Zhang Shaoqing, Ren Jian Sun (2015) Deep Residual Learning for

Image Recognition

[19] Sewade Ogun(2021) Consider using UAR instead of Accuracy for Imbalanced Classification tasks https://ogunlao.github.io/blog/2021/04/24/consider_uar_accuracy.html

[20] Powers, D. M. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. Journal of machine learning technologies, 2(1), 37-63.

圖片來源

[1] 高雄大學資工系教授 潘欣泰教授 上課用圖片

[2] Roger Jang (張智星) <http://mirllab.org/jang/books/audiosignalprocessing/overview-acousticFeatureStore.asp?title=22-2%20%AD%B5%B0T%AA%BA%B0%F2%A5%BB%C1n%BE%C7%AFS%BCx>