**研究計劃書**

**申請人：**

目 錄

[第 1 章 研究動機與目的 3](#_Toc181258395)

[1.1 研究動機 3](#_Toc181258396)

[1.2 研究目的 3](#_Toc181258397)

[第 2 章 文獻探討 5](#_Toc181258398)

[2.1 人工智慧在聲音辨識之應用 5](#_Toc181258399)

[2.2 聲音特徵提取 8](#_Toc181258400)

[2.3 深度學習(Deep Learning) 13](#_Toc181258401)

[第 3 章 研究內容、方法與工作項目 20](#_Toc181258402)

[3.1 研究方法設計 20](#_Toc181258403)

[3.2 使用工具 21](#_Toc181258404)

[第 4 章 預期成果 22](#_Toc181258405)

[第5章 未來展望：醫療聲音辨識處理系統 27](#_Toc181258406)

[參考文獻 30](#_Toc181258407)

# 研究動機與目的

## 研究動機

聲音辨識技術在醫療領域的研究動機和目的主要包括以下幾方面：首先，通過快速分析患者的語音、咳嗽聲和呼吸聲，提高診斷效率和準確性，特別是在面對大量患者時。其次，減少醫療資源負擔，利用自動化和遠程醫療減輕醫護人員工作壓力。再次，提升病患體驗，通過便捷的語音操作減少等待時間和不便。語音辨識技術還能增強疾病監測和管理，對慢性病和需要長期監測的病人提供持續的症狀監測。特別是在疫情期間，無接觸的篩查和診斷降低了病毒傳播風險。研究目的包括開發高效診斷工具、提升遠程醫療技術、改善患者與醫護人員的互動、推動醫療數據分析和管理、促進個性化醫療以及支持公共衛生和流行病學研究。這些目標旨在利用先進技術提升醫療服務效率、質量和覆蓋範圍，最終改善患者的健康和福祉。

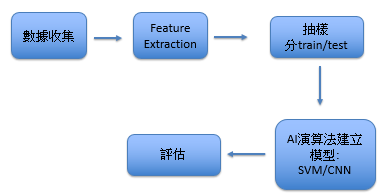
## 研究目的

聲音辨識技術在醫療領域的研究目的包括以下幾個方面：首先，開發高效的診斷工具，通過快速分析和識別疾病（如COVID-19、哮喘、肺炎等）的特徵聲音，提供輔助診斷支持，提高診斷效率和準確性。其次，提升遠程醫療技術，開發基於語音辨識的遠程醫療應用，讓患者在家中或遠程地點獲得高質量的醫療咨詢和診斷服務，減少面對面接觸，降低病毒傳播風險。此外，改善患者與醫護人員的互動，創建智能化的醫療語音助手，理解和回應患者需求，提供自然的互動體驗。

# 研究內容、方法與工作項目

## 研究方法設計

本研究參考文獻中的做法，採納MFCC或NMF做數據的預處理，將聲音訊號換算成向量後，再分割成訓練及測試集資料，再帶入不同模型，以準確率來評估模型優劣。使用之數據來源為https://www.kaggle.com/datasets/lynnkong/covid19voicediagnosis-2/data，該網站提供很多公開且珍貴之數據，並且其數據有一定之根據性及可靠度。



**圖 1研究流程示意圖**

在訓練過程中會使用測試資料測試模型是否過擬合狀況發生，若有則會用dropout方式關閉部份節點，因為數據有咳嗽聲以及呼吸聲，預計會以其中一個數據集建立模型後，再以這份模型去套用在另一份數據集，試著不同參數例如學習率測試是否可以直接套用或是可以用較少的步驟訓練，達到遷移學習的效果。

## 使用工具

本研究使用google提供的colab之python編譯器，做深度學習演練。

系統主要分三個架構：

1. 數據前處理

將聲音訊號變成數字資訊。

1. 模型建置模組

利用不同深度學習架構

1. 結果輸出模組

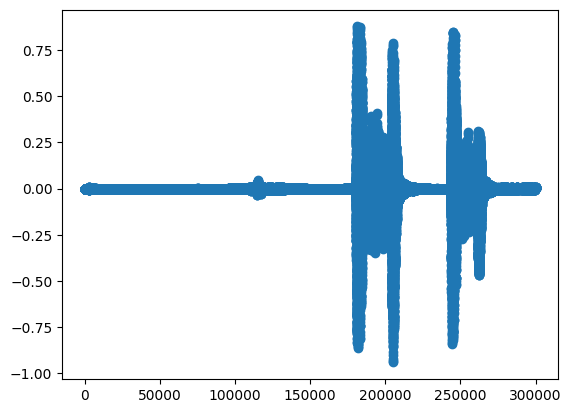
將程式寫成一份exe檔案，可以批次預測並輸出成一份html檔案。

# 成果展示

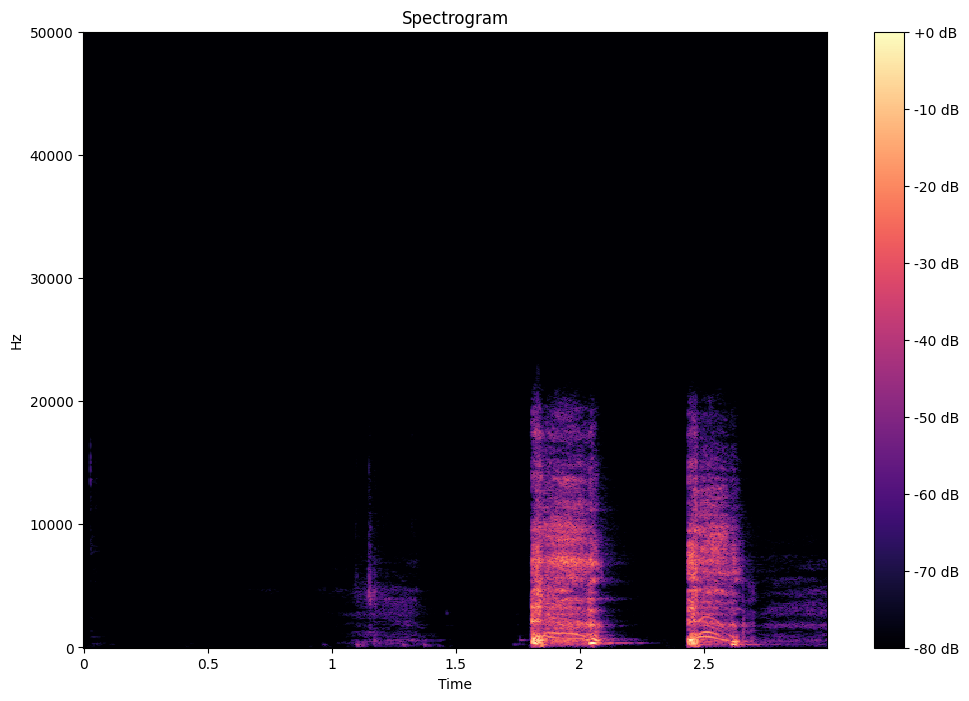
本研究得到下列之成果：

1. 可視化

原始的聲音若畫成波形如下，但這樣的數據顯然不能應用於機器學習或深度學習，因此本研究利用MFCC方法將一段聲音轉化成特徵向量或是用傅立葉轉換變成圖片如圖 3。



**圖 2 聲音波形圖**



**圖 3 聲音經過傅立葉轉換，X軸為時間**

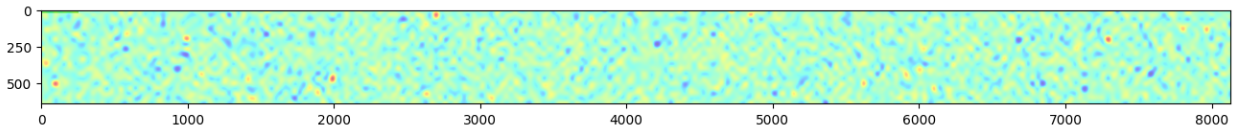
1. 與傳統之聲音辨識比較

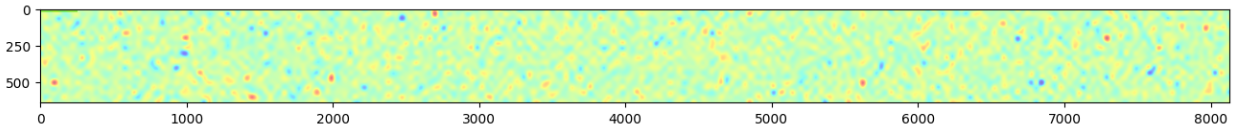
本研究採用MFCC或NMF萃取出該聲音之特徵，並且將這些特徵使用深度學習來訓練，準確率能達到90%以上或更高，希望未來該程式可以真正應用在醫療領域。

|  |  |  |  |
| --- | --- | --- | --- |
| 特徵萃取方法 | NMF | MFCC(取10筆) | MFCC(取20筆) |
| 使用SVM預測  準確率 | 70% | 90% | 91% |
| 使用CNN預測  準確率 | 69% | 92% | 93% |

1. 分析陽性與陰性聲音是否有特徵

原始檔之聲音包含呼吸及咳嗽如下圖，若人為去聽來分類，非常曠日廢時，並且需要很專注反覆地聽取才能辨析其中差異，本研究將聲音訊號數位化，嘗試尋找在某些頻率上或繪製成頻譜圖後可以找到一些特徵，讓我們了解是什麼樣之差異來辨識出陽性與陰性。

**圖 4 陽性的特徵圖**

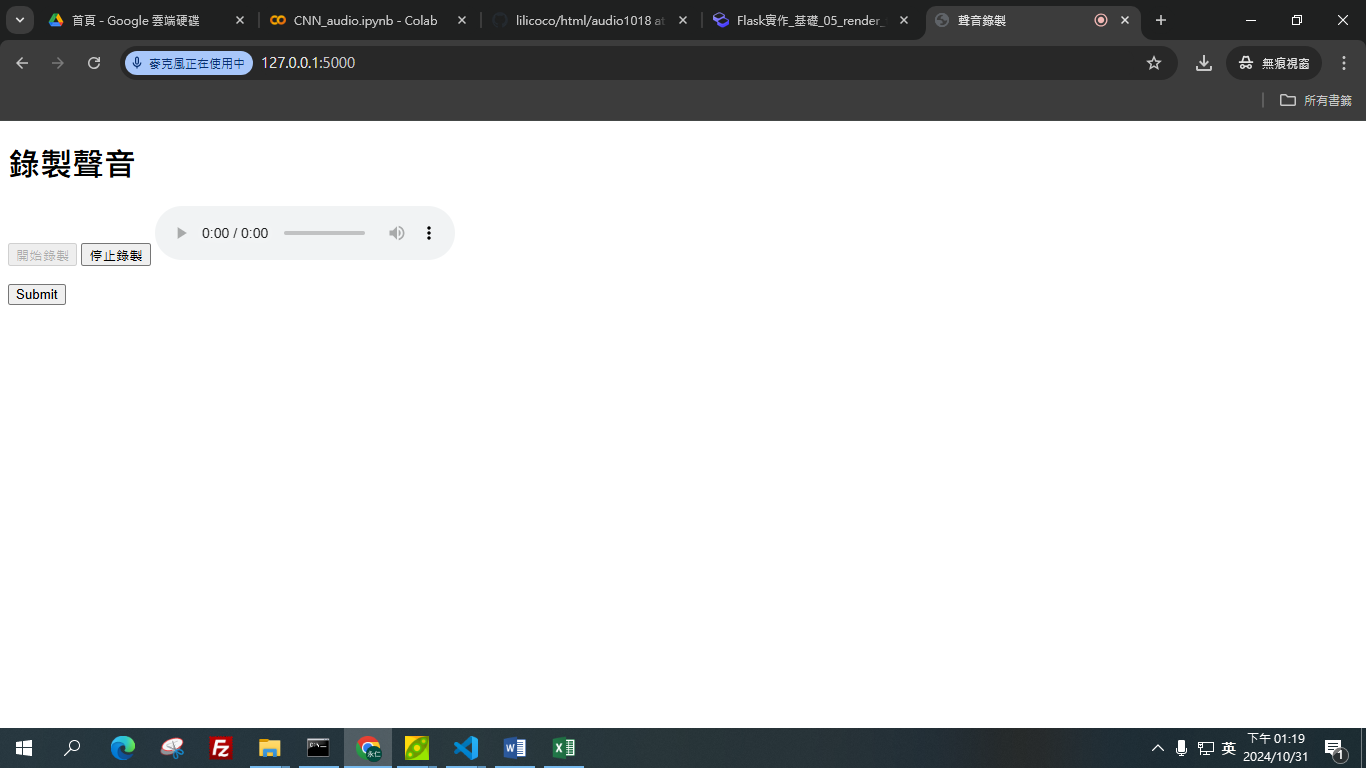


**圖 5 陰性的特徵圖**

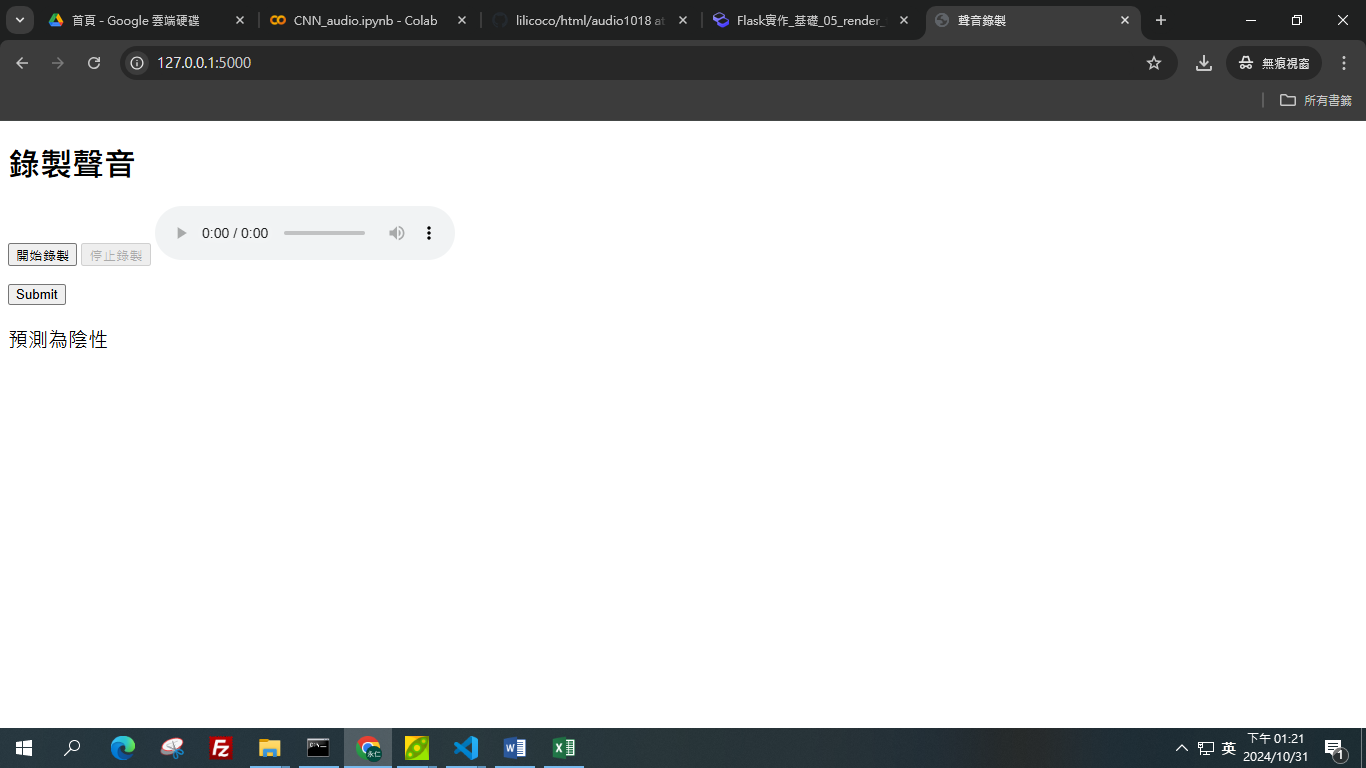
圖 4是透過擷取類神經網路內部最後一層的輸出的權重繪成熱圖，圖 5為陰性的圖，可以看出陽性的熱圖偏深色，因此會試圖將所以有陰陽性都繪製出熱圖，研究其中差異。

1. 製作成應用程式

本研究使用python中的flask套件製作成一網頁如圖 6，錄完聲音按下送出，後臺模型進行聲音預測，給予類別如圖 7。



**圖 6 網頁樣板**



**圖 7 網頁樣板**

# 第5章 未來展望：醫療聲音辨識處理系統

隨著技術的進步和醫療需求的增長，醫療聲音辨識處理系統在未來有著廣闊的發展前景。以下是對其未來展望的詳細分析：

(1) 提高系統精度

隨著更多醫療聲音數據的收集和精細標註，醫療聲音辨識系統的機器學習模型將持續優化，進而顯著提升系統的精度和穩定性。這些數據可以來自不同的醫療環境，如診所、醫院、遠程醫療平台等，涵蓋不同的病症和患者群體。特別是利用深度學習中的先進算法，如卷積神經網絡（CNN）和長短期記憶網絡（LSTM），能夠更有效地從複雜的聲音信號中提取特徵，提高疾病識別的準確率。此外，多模態數據融合技術（如結合聲音、影像和文本數據）將進一步增強系統的綜合判斷能力。

(2) 強化實時處理能力

為了滿足即時診斷和決策的需求，未來將引入高效的硬件加速技術，如使用GPU（圖形處理器）或專用的AI加速芯片，顯著提高系統的實時處理能力。這將使醫療聲音辨識系統能夠在診療過程中快速響應，提供即時的診斷建議和警示，特別是在急診室和手術室等需要快速決策的環境中。此外，邊緣計算技術的應用將使得系統能夠在設備端進行部分處理，減少延遲，提升患者體驗。

# 參考文獻

[1] 劉佳鑫。「基於深層類神經網路之 聲音事件偵測系統」。碩士論文，國立臺北科技大學電子工程系研究所。

[2] 蔡佩樺。「機器學習聲音辨識應用於車禍偵測之研究」。碩士論文，逢甲大學資訊電機工程碩士在職學位學程

[3] 張永霖。「基於卷積神經網路的聲音分類機制」。碩士論文，朝陽科技大學資訊與通訊系，2021。

[4] M. A. Hossan, S. Memon and M. A. Gregory, "A novel approach for FCC feature extraction," The 4th International Conference on Signal Processing and Communication Systems, pp. 1-5, Gold Coast, QLD, Australia, 13-15 Dec. 2010.

[5] 林蓬榮, “規劃網站品牌的競爭策略,” 管理雜誌, vol. 326, pp. 127-129, 民90.

[6] M. Neely, “Optimal pricing in a free market wireless network,” Wireless Networks (10220038), vol. 15, no. 7, pp. 901-915, 2009.

[7] 吳克振, “非優質品牌應該如何應變,” 設計, vol. 99, pp. 110, 民90，6月.

[8] T. Wasserman, “Zsystems' Linux System Cracks $199 Price Tag,” Brandweek, vol. 41, pp. 20, 2000.

[9] T. Wasserman, “Zsystems' Linux System Cracks $199 Price Tag,” Brandweek, vol. 41, no. 18, pp. 20, 2000.