國立台北大學資訊工程學系專題報告

Mobile Application Based on Deep Learning for Testing Covid-19 by Analyzing Cough Audio with BLE Stethoscope

專題組員: 曾柏瑋、謝秉諺、梁禎智、李秉倫

專題編號:PRJ- NTPUCSIE-111-006

執行期間:111 年 09 月 至 112 年 06 月

1. 摘要

近年來，隨著新冠肺炎的肆虐，全球有超過 7 億起確診案例，其中有很大一部份是因為疫情初期醫療資源與人力的匱乏，因此我們意識到在如此困難的情況下有效配置醫療資源和快速診斷的重要性。為了滿足這一需求，我們利用藍芽聽診器以及深度學習的技術，開發出能無線錄製音訊，並且用於疾病分析的 APP。

為了實現這個 APP,我們利用 Vision Transformer 和 MFCC 訓練出能分類 Covid-19 患者咳嗽聲的模型，再將訓練好的模型轉換成 TensorFlow Lite 模型，以部署在手機上，實現了在手機進行疾病分類的功能。希望運用本系統能有效解決醫療資源不足的問題，並且加速診斷流程。

2. 簡介

經過這幾年疫情的肆虐，我們目睹了過去醫療資源嚴重不足的情況，並深刻體會到醫療資源匱乏所帶來的後果，也了解到醫療資源的重要性，同時我們也見證了人工智慧隨著時代的發展，快速成長到能夠解決各種不同領域的問題，應用於人們的日常生活中。

在過去的研究中，已經發現人體在感染某些疾病時，會導致心肺音、呼吸聲或者咳嗽聲產生變化，這些聲音被視為是一種判斷疾病的依據，此外，也有研究嘗試提取這些聲音的特徵，使用機器學習或者深度學習的方式來進行疾病的檢測，並取得了良好的成果。

因此，我們決定開發一套系統，能夠錄製使用者的聲音，並結合深度學習的技術，對錄製的聲音進行分類，來區分使用者是否感染疾病。

在考慮到疫情的嚴重性和成果的實用度，我們將 COVID-19 作為本系統的研究目標，並使用咳嗽聲做為分析的資料，此外，在考慮到 COVID-19 具有高傳染性的特性，我們選擇結合 Bluetooth Low Energy(BLE)，讓使用者可以遠距離錄製聲音，降低傳染的風險，再者，考量到系統的便利性，我們決定開發一款手機應用程式，並將模型部屬於手機應用程式，節省架設伺服器和網路傳輸的成本，同時也保障了使用者的隱私。

3. 專題進行方式

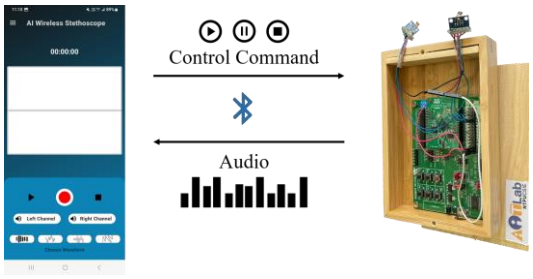
我們將使用 Android Studio 做為開發工具,設計一個 Android 手機應用程式，搭配 BLE 無線聽診器(RTL8762DW 開發板)，透過無線傳輸實現即時聲音錄製。此外，我們將在電腦上使用 PyCharm 開發環境預先訓

練 COVID-19 患者咳嗽聲分類模型，並將該模型轉換成 APP 所支援的格式並部署到 APP 中。

本系統可分為主要三大項目，無線即時聲音錄製、AI 疾病分類以及使用者介面設計。

3.1 無線即時聲音錄製

本系統將利用手機以 APP 與聽診器藉由低功耗藍牙來建立無線連線，並由 APP 傳送控制訊號來控制聽診器進行錄音，當聽診器收到錄製聲音的訊號後，會將聲音資料轉換為 PCM 格式，並採取 SBC 編碼方式進行壓縮，減少需要傳輸的資料量以達到即時錄製的效果，接著將數個經 SBC 編碼後的訊框加上 Header 後做成封包傳送到手機端，當 APP 收到由聽診器所傳送的封包後，需要根據該封包的 Header 進行解封，再按照訊框中的格式獲取 SBC 編碼時所用的參數，使用這些參數對經過編碼的音訊進行解碼以得到音訊資料，並即時在手機上撥放音訊與繪製聲波圖，同時在手機上建立暫存檔，將音訊資料寫入暫存檔，以利後續進行撥放與歸檔。



圖一 手機與聽診器無線傳輸示意圖

3.2 AI 疾病分類

在錄音完成後，使用者可以使用預先訓練好且部屬於手機上的模型來對已經錄製完成的聲音進行分析。

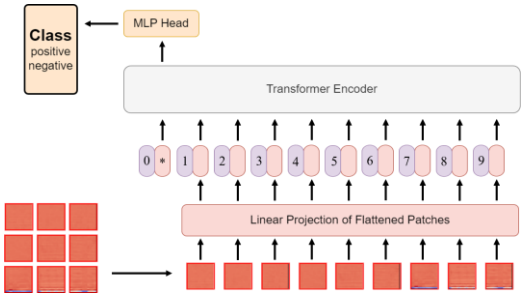
為了達到上述的目標，我們將分成五個步驟執行，分別是資料的預處理、特徵的提取、訓練模型、轉換成手機模型以及部屬模型。

我們所採用的資料集是由劍橋大學使用群眾外包的方式取得的資料集，資料集中包含兩類具有標籤的音訊檔案，分別是屬於 COVID-19 確診患者的咳嗽聲以及不屬於 COVID-19 確診患者的咳嗽聲，由於每個音訊檔案的取樣率、長短及品質各不相同，因此在進行資料的預處理時，我們需要先將音訊檔案進行重新取樣，並且將音訊檔案的長度截斷為兩秒，以確保資料的一致性。

接著在特徵提取的部分，我們將原本的聲音數據轉換為 MFCC，MFCC 具有以下三大優點：抗噪能力、特徵壓縮和序列性特徵，由於我們資料的品質不一，因此使用具有抗噪能力的 MFCC 做為輸入有助於提升模型的性能，特徵壓縮則大大減少了我們計算和記憶的需求，加快模型的訓練和收斂速度，同時也可防止模型過擬合，最後，MFCC 保留了資料的時序訊息，因此透過提取聲音資料的 MFCC，我們能夠很好得捕捉到聲音的特徵，並用於後續的分析和模型訓練。

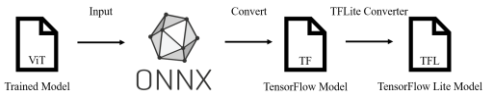
模型部份我們使用 pretrained 的 Vision Transformer 模型進行微調，並添加全連接層以實現二元分類，使用 pretrained 的模型是希望能夠縮短訓練模型所花費的時間，同時能得到更好的結果，我們所使用的 Vision Transformer 模型的輸入為 384x384 像素的影像，並且會將輸入的影像分割成數個 16x16 像素的區塊，這些區塊經過線性嵌入並添加位置資訊後會變成一個序列，最後再送到複數層的 Transformer Encoder，這些 Transformer

Encoder 能夠幫我們提取更高層次的特徵和學習更複雜的圖像，最後經由全連接層輸出二元分類的結果，採用 Vision Transformer 的優點是因為 Vision Transformer 可以捕捉影像中不同區塊之間的位置關係，這對於我們所使用的帶有時間序列關係的聲音資料相當有效。



圖二 ViT 模型架構

模型訓練完成後，我們需要將模型轉換成可以在手機上直接使用的格式，首先將模型轉換成 Open Neural Network Exchange 格式，再轉換成 TensorFlow 模型，最後使用 TensorFlow Lite Converter 將模型轉換成 TensorFlow Lite 格式，方能在手機 APP 中使用。



圖三 模型轉換流程

最後在手機 APP 中部屬模型，透過將已經轉換成 TensorFlow Lite 格式的模型作為 Asset 加入 Android Studio 的專案中，即可在 APP 中使用，不使用 Android Studio 本身所具有的匯入 TensorFlow Lite 模型的功能的原因是這種做法可以不用壓縮模型的大小，以避免因壓縮模型而導致模型的準確率下降。

3. 3 使用者介面設計

使用者介面將影響使用者的使用意願及系統操作上的便利性，是每個完整的系統都應該特別留意的部分，為了設計出良好的使用者介面以及簡潔精美的畫面，讓使用者能夠一目了然，提供使用者良好的使用體驗以及減少操作時的疑惑，我們將參考 Material Design 的內容進行設定和規劃，並融入自身設計風格和使用經驗，嘗試提供使用者直覺且具有彈性的使用者介面與使用體驗。

4. 主要成果與評估

我們在 APP 中實現了 SBC 解碼器部分，完成了無線即時聲音錄製的功能，此外還包含了以下其他特殊的功能

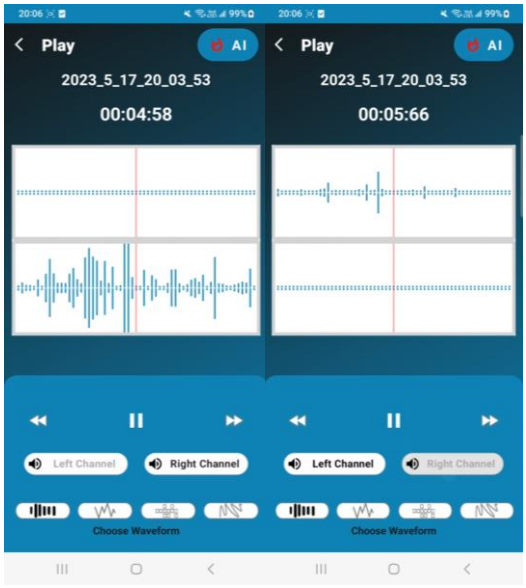
4. 1 錄音操作：

使用者可以隨時暫停錄音，並且可以從暫停的地方繼續錄音，在暫停錄音時，使用者還可以拖曳聲波圖以檢視過去的錄音，並且選擇任意時間點進行撥放，提供使用者更高的靈活性和控制權。

圖四 錄音介面

4.2 切換左右聲道：

使用者在錄製聲音時可以個別選擇是否撥放左、右聲道，透過切換左右聲道，我們可以去比對不同聲道間的差異。



圖五 左右聲道切換

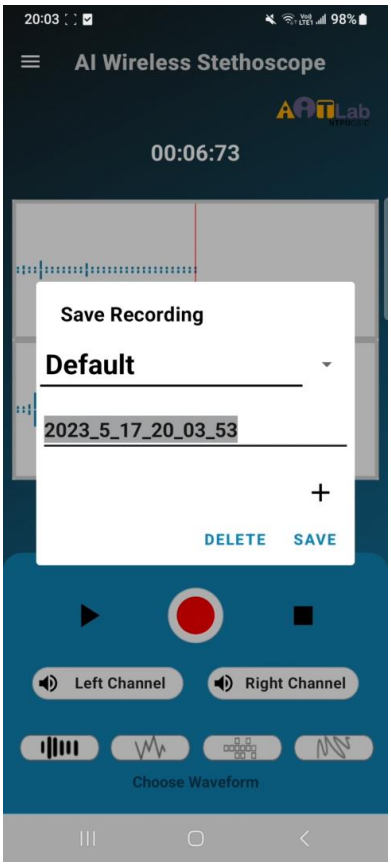
4.3 波形顯示：

我們提供四種波形來顯示聲音的即時動態，使用者可以根據場景、需求或是個人喜好迅速切換想要顯示的波形，提供多角度的觀察，獲取更全面的數據。

圖六 四種波形顯示畫面

4.4 檔案管理：

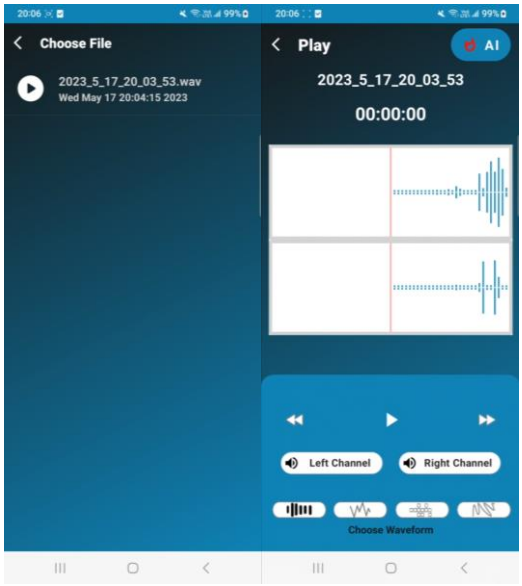
在錄音結束後，使用者可以新增和自選資料夾並且可以更改錄音檔的名稱，方便對檔案進行分類和管理。



圖七 存檔畫面

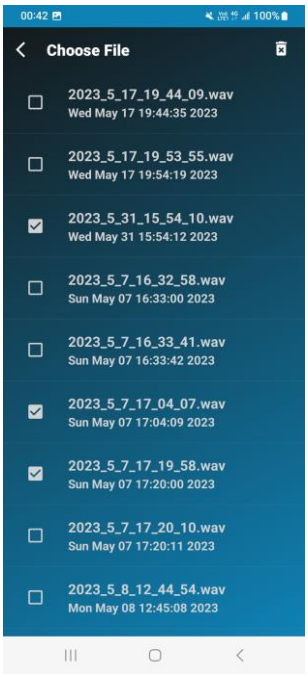
音訊檔案將會以 WAV 檔的形式儲

存，檔案儲存完畢後，使用者可以在歷史資料中看到過去所儲存的錄音檔，並且可以撥放儲存的錄音檔或是進行 AI 疾病分類，方便使用者回顧。



圖八 歷史資料清單與播放畫面

使用者還可以對資料夾或者錄音檔進行管理，只需長按，即可切換到刪除的功能，能一次選取多個資料夾或檔案進行刪除，方便對檔案進行分類和管理。



圖九 檔案刪除功能

4.5 AI 疾病分類：

我們最終所訓練出來的模型其準確率為 77%，F1 score 則為 0.7。而在速度方面，我們的模型在當前中階手機上測試只需數秒即可完成分類。

AI Model	Accuracy	F1 score	Model size
Vision Transformer	77%	0.7	436MB

圖十 模型總結

4.6 通知功能：

由於儲存錄音需要花費一點時間，為了能夠讓使用者在這段時間內可以進行其他操作，我們使用通知功能來告訴使用者目前哪些檔案正在儲存以及哪些檔案已經儲存完畢，使用者也可以在設定中關閉通知功能，讓使用者能夠依據個人需求進行調整。

圖十一 錄音完成通知畫面

4.7 語言切換：

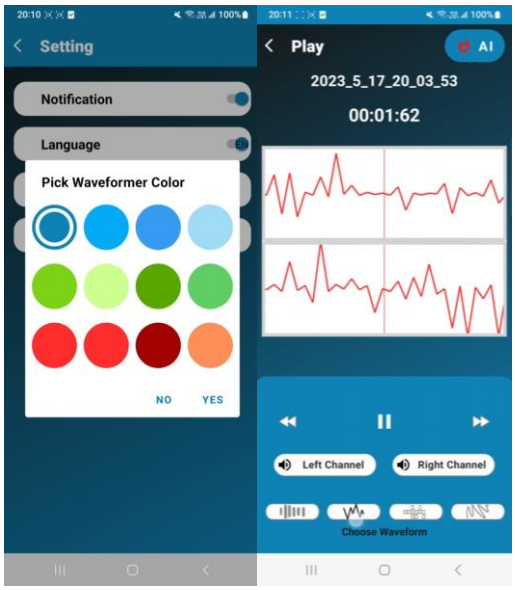
我們提供中文及英文供使用者進

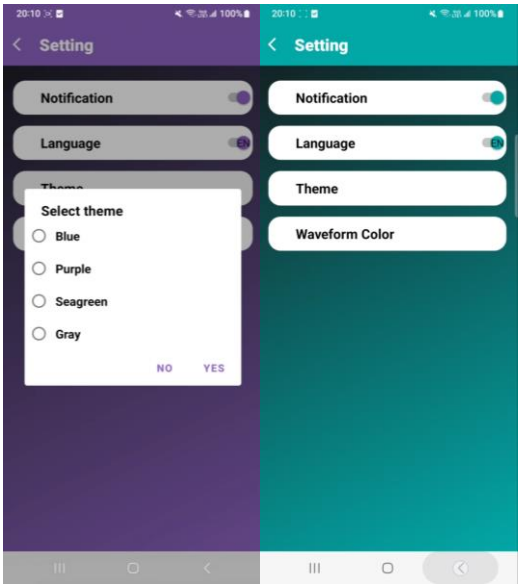
行切換，使用者可以根據自身習慣選取合適的語言，讓使用者可以更方便去進行操作，並且擴大我們的用戶範圍。

圖十二 中英文介面

4.8 自訂主題與波形顏色：

通過自定義波形顏色和主題，使用者可以根據自己的喜好和偏好的來個性化 APP 的外觀。不同的主題和顏色組合可以為用戶帶著獨特的視覺感受，帶給使用者舒適的操作體驗。





圖十二 使用者自訂介面功能

5. 結語與展望

我們成功使用 Vision Transformer 模型、Android 手機 APP 搭配 RTL-8762DW 開發版，完成無線聽診器與 AI 疾病分類的系統，透過這項研究，我們希望可以解決醫療資源匱乏的問題以及減少病患與醫生的接觸，未來我們期望能夠提升聽診器的性能，讓聽診器能收到更多種類的生理音訊，並縮小聽診器的體積，以方便攜帶與使用，此外，我們也期望能結合與訓練其他種類的疾病分類模型，增加本系統的應用範圍，以及改進 UI/UX 的設計，降低操作的難度並改善使用者的體驗。

6. 銘謝

特別感謝指導教授提供研究上的幫助以及實驗室的研究資源，也感謝實驗室學長姊的協助，讓我們能夠克服一路上遭遇的種種困難，最後得以順利完成專題。

7. 參考文獻

[1] Bluetooth SIG, "Advanced Audio Distribution Profile 1.4", 2022

[2]C. Brown, J. Chauhan, A. Grammenos, J. Han, A. Hasthanasombat, D. Spathis, T. Xia,P. Cicuta and C. Mascolo, "Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data", preprint, arXiv:2006.05919 [cs.SD] (2020).

[3] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit and N. Houlsby, "An image is worth 16 x 16 words: Transformers for image recognition at scale", preprint, arXiv:2010.11929 [cs.CV] (2020).

[4] T. Hoang, L. Pham, D. Ngo and H. D. Nguyen, "A Cough-based deep learning framework for detecting COVID-19," 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Glasgow, Scotland, United Kingdom, 2022, pp. 3422-3425, doi: 10.1109/EMBC48229.2022.9871179.

[5] L. Khriji, A. Ammari, S. Messaoud, S. Bouaafia, A. Maraoui and M. Machhout, "COVID-19 Recognition Based on Patient's Coughing and Breathing Patterns Analysis: Deep Learning Approach," 2021 29th Conference of Open Innovations Association (FRUCT), Tampere, Finland, 2021, pp. 185-191, doi: 10.23919/FRUCT52173.2021.9435454.