

以深度學習與智慧問診系統為基礎之 AR 穴位互動系統

魏仲彥、徐伯元、莊汶娟、李偉德、鄭智陽、趙于翔

國立金門大資訊工程學系

*yxzhao@nqu.edu.tw

論文摘要

按穴道是保養身體的方法之一，有很多能夠治癒疾病的特效穴位，不但可以達到治癒單一疾病的效果，還可以調整全身機能，強身健體，但是穴位遍布人體，穴道的實際位置往往要藉由書本或他人的協助，初學者才能得知，我們希望使用者能一眼看出手部穴道在自己身上的位置與作用，使得使用者能更容易運用手上以及頭部的穴道，做到即時緩解症狀的功效，進一步預防疾病的發生。本系統一共分為三個部分，軟體端、手機端、伺服器端。軟體端使用 Unity 進行開發，使用者只需要使用手勢，就可以操控軟體。手機端可以傳送類別訊息給軟體端，讓螢幕顯示對應的疾病穴位。手機端也可以與伺服器連線，讓使用者進行 AI 問診。本計劃的創新點在於結合 Unity、手機 APP 和 AI+AR 技術，可以透過 AR 互動，點擊穴位，顯示出穴位資訊，有別於傳統的書本教學。

關鍵詞：穴道定位、深度學習、手勢辨識、AI 問診、AR 穴位互動學習

前言

目前人工智慧在自然語言處理、影像辨識、視覺檢索、醫療影像分析等領域，已有了很大的進步 [1, 2]，尤其是最近幾年來，隨著電腦規格的快速提升，並且各領域人才的成長，人工智慧的實力也在逐年增強。本論文結合人工智慧、增強現實技術和手機應用程式的穴位按摩系統。我們希望透過這樣的系統，能夠讓人們更容易學習穴位知識，並且可以自行進行緩解身體不適的按摩治療，提高健康水平。本計畫所要探討研究問題包含：

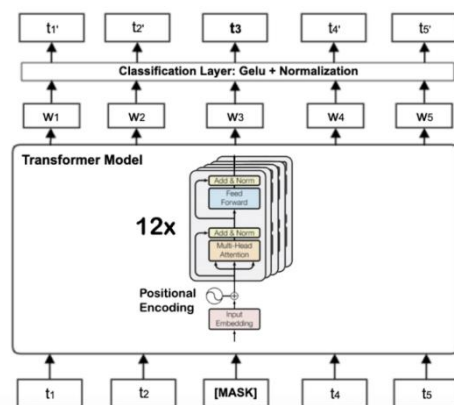
- (1) 如何讓穴位正確的顯示在螢幕上？
- (2) 如何運用神經網路進行手勢辨識？
- (3) BERT 模型進行疾病對話推斷的效果如何？
- (4) 對於疾病分析技術的改善與發展？

文獻探討

穴位是中醫保健的重要元素，近年來人工智慧技術在穴位辨識上有了長足的進步。早期使用 CNN 模型和數學運算[3]或是邊緣偵測[4]推斷穴位，而現在可以取得人體座標點的模型，如 Mediapipe[5]。過去有論文使用 3DMM 模型來達到臉部穴位的顯示[6]，但其複雜的計算需求限制了它實時運行的可能性。Mediapipe Face Mesh 是一個實時面部標誌檢測系統，使用機器學習來預測面部 468 個關鍵點的 3D 位置，相比 3DMM 模型，

Mediapipe 效能更好，適用於各種設備上運行。近年也有研究使用 Mediapipe 做臉部穴位顯示[7]，但功能較為單一。本論文在 Mediapipe 基礎上進一步開發，實現了不僅有臉部穴位顯示，還包括手部穴位、穴位分類、手勢互動等功能，讓使用者可以有更完善的體驗。

在 AI 問診中使用 Rasa[8]+BERT[9]，Rasa 是一個開源的對話人工智慧平台，主要用於開發和建立智能對話機器人。Rasa 的核心功能是理解使用者輸入的意圖和實體，然後進行相應的回應。Rasa 的訓練資料由自然語言處理 (NLP) 模型來處理。當使用者輸入文字時，Rasa 會根據使用者的輸入，利用已經訓練好的模型來預測使用者的意圖，然後進行相應的回應。而 BERT 也是用於自然語言處理 (NLP) 的預訓練深度學習模型，由 Google 研發出來的 Transforms[10] decoder 部分延伸出來的。BERT 能夠對句子進行雙向編碼，也就是同時考慮句子中所有詞的前後文脈信息。這使得 BERT 在許多 NLP 任務中都有出色的表現，包括情感分析、機器翻譯、關鍵字提取和問答系統等。本論文使用的 BERT 的架構包括了 12 個層 (見圖一)。BERT 在訓練過程中使用了一種稱為自適應學習的技巧，使得模型能夠根據輸入的句子動態地適應不同的輸入長度。此外，BERT 也使用了殘差連接和自注意力機制 (self-attention)[10]，這使得模型能夠更好地捕捉句子中的相關性。透過 BERT 模型，我們希望能夠藉由與用戶的對話來推斷用戶的疾病。



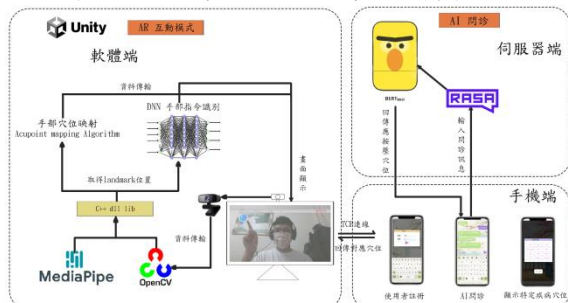
圖一、BERT 架構圖

研究方法

1、系統架構

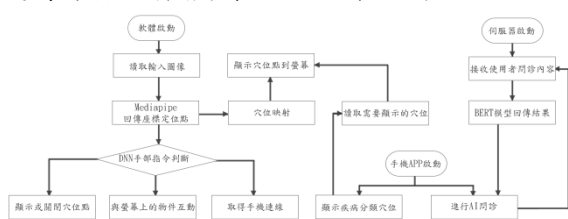
圖二是我們的系統架構圖，軟體端使用 Unity 開發，並使用 MediaPipe 來獲取人體座標點，利用演算法可以從 MediaPipe 得到的座標點映射到頭部和手部的穴位，螢幕可以顯示出穴位，讓使用者

能夠學習和了解穴位的知識。本論文還設有一個 Android APP 讓使用者下載，可以透過這個 APP 進行 AI 問診，並記錄使用者的身體狀況和相關資訊。使用者在問診時，APP 能夠連到伺服器端進行 AI 問診，伺服器會進行自然語言理解 (NLU) 並判斷使用者的身體狀況，回傳手機螢幕上顯示。



圖二、系統架構圖

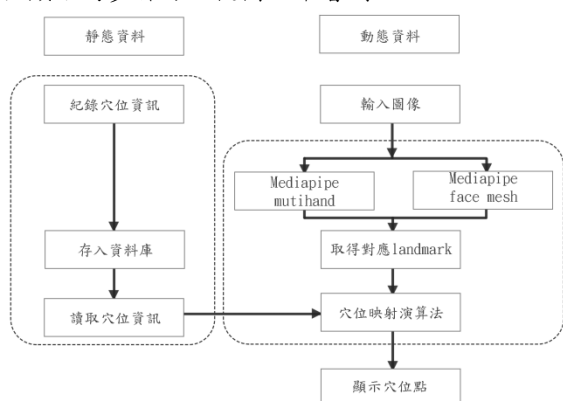
在研究步驟中，我們主要分成三大部分來討論，分別為軟體端、手機端以及伺服器端，可以從圖三中看到，軟體端負責處理運算邏輯、AR 穴位顯示邏輯、手部姿勢辨別，伺服器端負責 AI 問診，而手機端則是顯示接收伺服器的訊息，以及傳送需要顯示的疾病特效穴位給軟體端。



圖三、系統流程圖

2、穴位映射

我們的手部定位以及頭部定位是利用資料庫的位置資料，裡面的資訊包括參考 Meidapipe 座標位置、與座標位置的相對距離，以及穴位名稱，穴位顯示的步驟可以從圖四中看到。



圖四、穴位映射流程圖

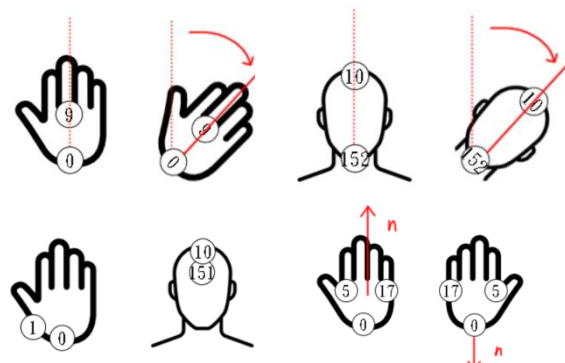
頭部和手部穴位旋轉本論文使用公式(1)進行判斷，利用最上面的座標點與最下面的座標點當作基軸(見圖五(a)、(b))，當手部進行旋轉時，會根據基軸的轉動量生成轉移矩陣，如果公式(2)，讓其

他穴位可以根據這個轉移矩陣進行旋轉，精確的映射到對應的穴位上。我們還根據不同人的骨骼，對穴位距離進行調整(見圖五左下(c))

$$\varphi = \frac{\vec{v}_{0to9} \cdot \vec{nv}_{0to9}}{\sqrt{(x_0-x_9)^2+(y_0-y_9)^2} \times \sqrt{(nx_0-nx_9)^2+(ny_0-ny_9)^2}} \quad (1)$$

$$\text{where } \vec{v}_{0to9} = (x_0 - x_9, y_0 - y_9), \\ \vec{nv}_{0to9} = (nx_0 - nx_9, ny_0 - ny_9)$$

$$[\alpha, \beta] = [x, y] \cdot \begin{bmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{bmatrix} \times z \quad (2)$$



圖五、穴位定位判斷示意圖：

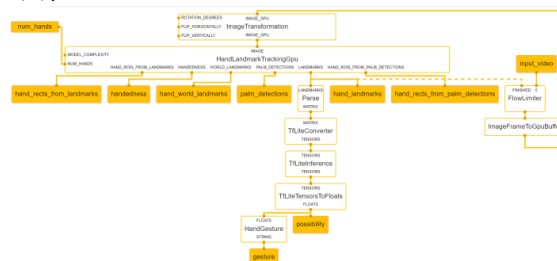
- (a)頭部選轉判斷；(b)手部選轉判斷；
(c)手部與頭部基本向量；(d)手部正反判斷

在人體手部正反面，我們為了要顯示不同的穴位，我們取手掌底部、拇指、小拇指三個點建立兩個向量(見圖五右下(d))，並使用公式(3)對它們取垂直法向量，用來判斷手部的正面和反面

$$n = \vec{v}_{0to17} \times \vec{v}_{0to5} \quad (3)$$

3、DNN 手部辨識

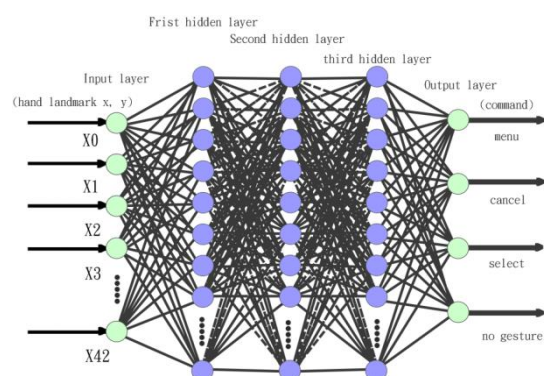
我們自定義 DNN(Deep Neural Network)神經網路的 TFLite 模型，並創立一個 Mediapipe Graph(如圖六)，其中有使用到手部辨識模型，我們根據 Mediapipe 模型回傳的定位點，判斷使用者手勢。



圖六、Mediapipe Graph HandGestureDetection

HaGRID(HAnd Gesture Recognition Image Dataset) [12]是一個手勢識別圖像數據集，本計畫使用其中的 10807 筆資料進行 DNN 模型的訓練。這個模型的輸入共有 42 個特徵，包括 Mediapipe 回傳的 21 個手部定位點的 X 和 Y

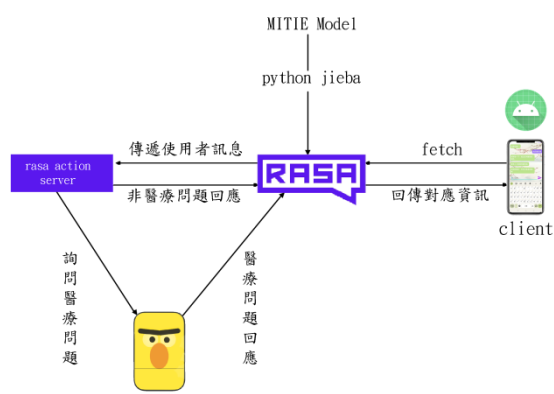
座標。因為輸入的特徵是座標位置，所以我們只需要三層 DNN 神經網路，就可以快速的分類出目前手勢是屬於哪一類型。我們的輸出一共有四類：呼叫選單、取消動作、選擇手勢以及無手勢。



圖七、DNN 手部辨識模型

4、BERT AI 問診

圖八展示了本計畫的 AI 問診架構，本計畫自行架設伺服器，並在伺服器上建設簡單的問診機器人。我們透過 Rasa 來建立聊天機器，當使用者的意圖為詢問疾病或是症狀，本計畫訓練的 BERT 模型就會根據使用者輸入，給出對應的特效穴位。



圖八、AI 問診架構圖

Rasa使用的NLP是使用MITIE模型來實現的。MITIE 模型可以將輸入文本轉換為向量表示，然後利用這些向量進行意圖和實體識別。因為本系統使用中文，所以需要先準備中文的詞向量模型，這裡採用了網路上的中文詞文本[11]。透過 jieba 進行分詞，可以將中文文本轉換為詞的序列。接著將分好詞的文本資料提供給 Rasa 進行模型訓練。在訓練完成後，當使用者輸入文字時，Rasa AI Server 會根據使用者的輸入，利用已經訓練好的模型來預測使用者的意圖，然後進行相應的回應。另外，當使用者輸入疾病相關的意圖時，本系統建立的 action server 會將使用者的輸入送至 BERT 模型進行處理。

本計畫使用了 BERT 的預訓練模型使用 bert-base-chinese[13]，該模型具有 12 層 layer、768

層 hidden、12 個 heads，參數規模達到 110MB。在進行模型訓練前，我們對 BERT 模型進行 token 的輸入初始化，並加入我們從維基百科蒐集的資料。目前我們的資料集中包含 109 種類型的疾病資料，以及與這些疾病可能發生的症狀相關的對話資訊（見表一）。總共有 623 筆資料。我們使用這些資料進行模型的 fine-tuning，即在 BERT 的最後一層接上一個新的簡單的分類器，用於識別使用者輸入的疾病類別。

disease	text
流行性感	發燒、咳嗽、喉嚨痛...
急性咽喉炎	咽部疼痛，吞嚥疼痛...
....
中暑	頭暈、頭痛、口渴...
消化不良	持續性食慾不振...

表一、BERT fine-tune data

研究結果

穴位映射效果展示

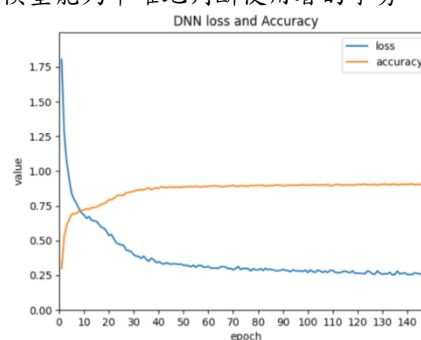
可以從圖九中看到，手部穴位貼合在固定的位置，並不會因為手部選轉或是其他姿勢而跑掉，這證明了我們的穴位映射演算法準確且有效。



圖九、穴位映射效果展示

DNN 模型效能評估

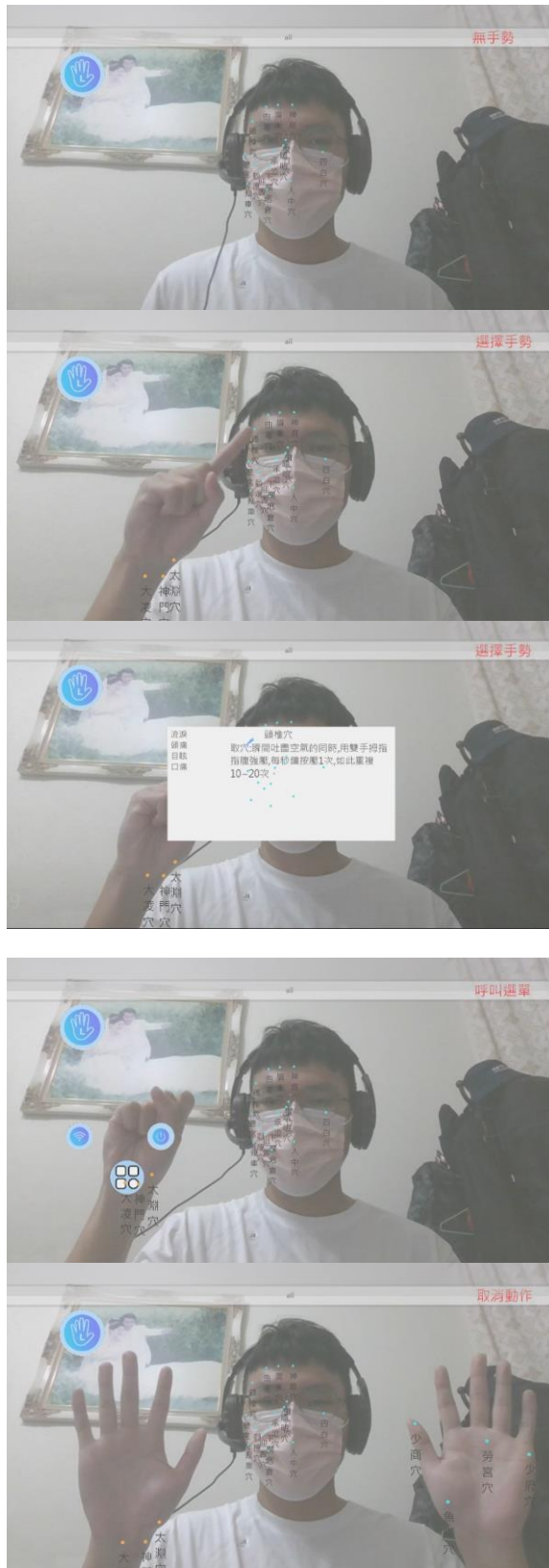
本論文使用 DNN 來進行手部辨識，我們使用的 epochs 為 150，batch size 為 16，透過訓練過程中的損失函數的觀察，我們可以在圖十中看到損失函數在訓練過程中慢慢下降，而準確率則慢慢上升，顯示模型的學習效果良好。本計畫的 DNN 手部姿勢辨識模型達到了高達 90% 的準確率。這表示我們的模型能夠準確地判斷使用者的手勢



圖十、DNN 模型準確率與損失函數

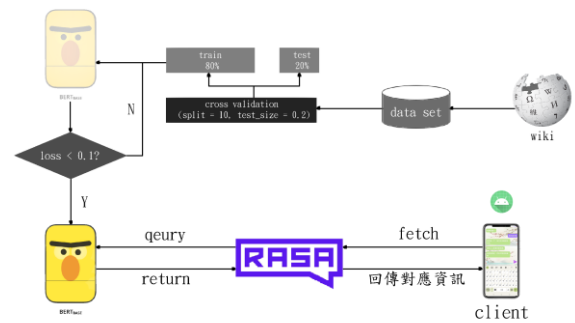
DNN 應用展示

以下是軟體端實際測試，可以看到穴位點貼合在人體上，並可以判斷手勢，選擇手勢可以與螢幕上的物件互動，呼叫選單可以控制手機連線和開關穴位點。

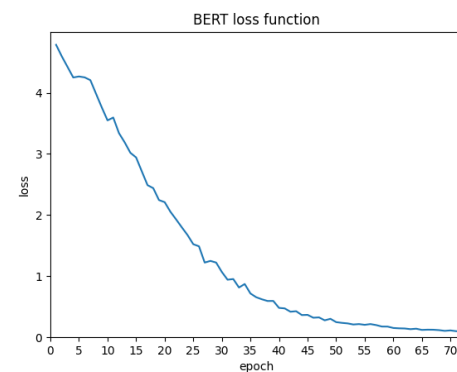


圖十一、DNN 手部辨識

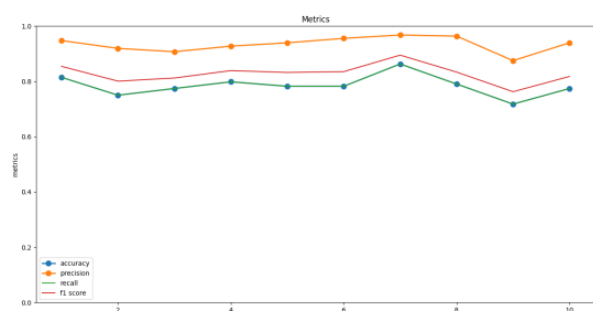
BERT 模型效能評估



在本論文中我們將資料集分為八成的訓練集和兩成的測試集對 BERT 模型進行 fine-tune，在訓練模型時，我們使用 epochs 為 72，batch size 為 10。透過觀察圖十二中的損失函數變化，我們可以看到損失函數在訓練過程中逐漸下降，顯示模型的學習效果逐漸改進。當損失函數小於 0.1 時，我們才結束訓練。每次訓練完畢後，我們將得到四個模型指標，包括 Accuracy、Precision、Recall、F1-Score。為了確保實驗可信度，我們使用 Cross Validation 進行十次的訓練並將十次的模型指標結果相加後取平均，作為最終的模型指標。這樣的做法有助於減少單次訓練的隨機性對結果的影響，並且能夠更全面地評估模型的性能。根據我們的結果，在預測疾病方面，我們的模型表現出色，達到近八成的準確率。這意味著我們的模型對於使用者輸入的症狀和疾病類別能夠給出準確的預測結果。圖十三中的結果進一步驗證我們訓練的 BERT 模型在問診機器人中的有效性和可靠性。



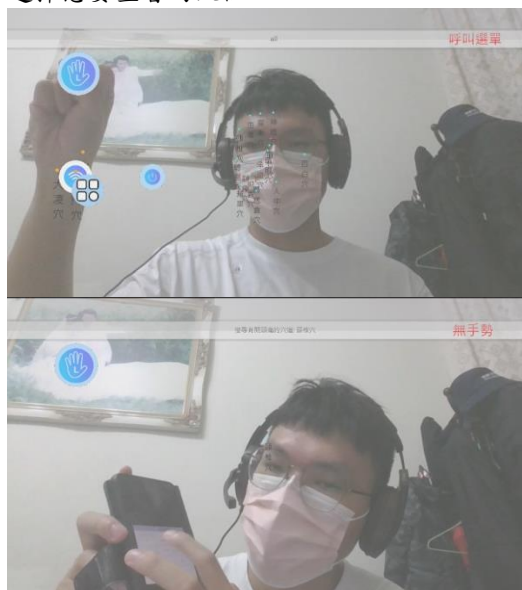
圖十二、BERT loss function



圖十三、BERT model metrics

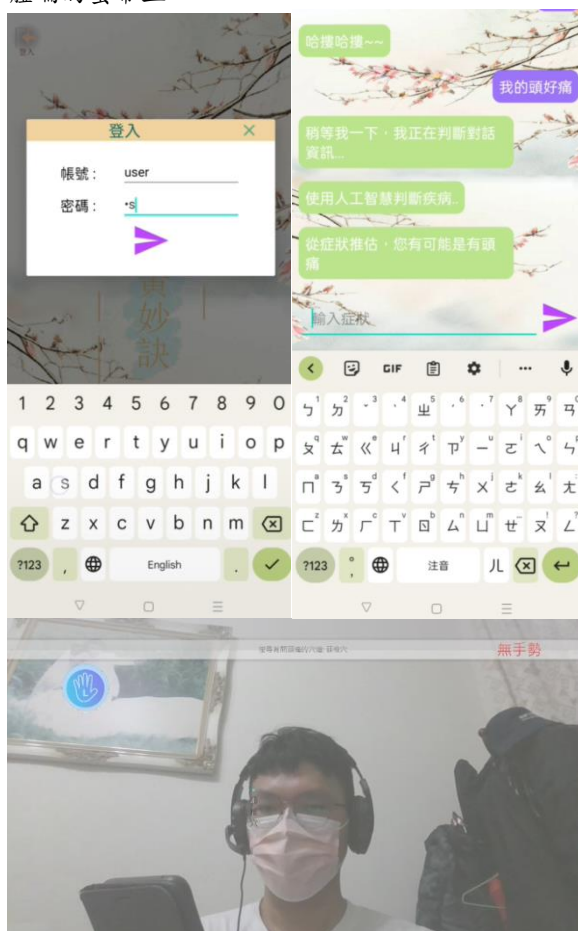
手機應用展示

當軟體與手機連線以後，可以直接使用 APP 選擇想要查看的穴位



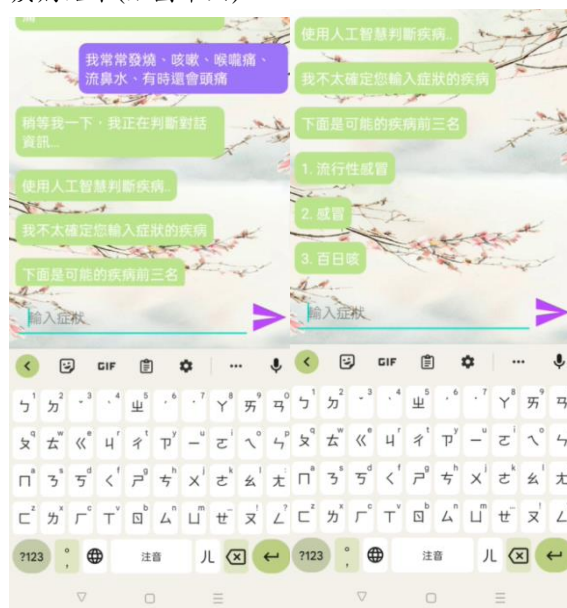
圖十四、穴位分類

手機登入後，可以選擇場景，下面選擇 AI 準診，當使用者輸入疾病相關的內容，BERT 會回傳可能的疾病，並顯示可以治療疾病的特效穴位到軟體端的螢幕上。



圖十五、BERT AI 問診

在 BERT 問診中，使用者如果輸入的醫療問題模糊不清，或是 BERT 推測的機率過低的話，我們會取機率最高的前三名，提供使用者查看有可能的疾病結果(如圖十六)。



圖十六、BERT AI 問診

結論與討論

本論文開發讓使用者可以以互動式的方式了解穴位，我們結合 MediaPipe 實現手勢控制軟體，並把穴位映射到螢幕上，並使用 Rasa+BERT 模型進行疾病對話推斷，未來，本系統可以進一步整合更多的診斷技術和健康管理功能，以提供更全面的健康管理方案給使用者。具體而言，本計畫之預期結果如下：

- (1) 建立演算法，讓穴位映射到對應的位置。
- (2) 使用 DNN 達到手部姿勢辨識的神經網路。
- (3) 提出一種基於 BERT 的疾病分析技術，討論本專題對於疾病推斷的準確度，和應用成果。
- (4) 本專題未來會加入更多模型訓練，並反饋到資料集，期望本專題的資料集可以延續使用，並讓系統可以造福更多使用者。

誌謝

本研討會要特別感謝澎湖科技大學的教師同仁大力協助，讓本研討會可持續舉辦。

參考文獻

- [1] Baohua Sun, Lin Yang, Wenhan Zhang, Patrick Dong, Charles Young, Jason Dong, Michael Lin, Demonstration of Applications in Computer Vision and NLP on Ultra Power-Efficient CNN Domain Specific Accelerator with 9.3 TOPS/Watt, 2019, pp. 1213-1223.
- [2] Andreas S. Panayides, Amir Amini, Nenad D. Filipovic, Ashish Sharma, Sotirios A. Tsaftaris, Alistair Young, David Foran, Nhan Do,

- Spyretta Golemati, Tahsin Kurc, Kun Huang, Konstantina S. Nikita, Ben P. Veasey, Michalis Zervakis, Joel H. Saltz, Constantinos S. Pattichis, AI in Medical Imaging Informatics: Current Challenges and Future Directions, 2020, pp.1-15.111-114, doi:10.1109/ICICCS.2016.7542339.
- [3] Gee-Sern Hsu, Hsiao-Chia Peng and Kai-Hsiang Chang, Landmark Based Facial Component Reconstruction for Recognition Across Pose, June 23-28 2014, pp. 34-39
 - [4] Menglong Chang, Qing Zhu, Automatic location of facial acupuncture-point based on facial feature points positioning April 1 2017, pp. 545-549
 - [5] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, Matthias Grundmann , MediaPipe: A Framework for Building Perception Pipelines, Jun 14 2019, pp.1-8.
 - [6] Yi-Zhang Chen , Corky Maigre , Min-Chun Hu and Kun-chan Lan, Localization of Acupoints using Augmented Reality, June 20 2017, pp. 239-241
 - [7] Menghe Zhang , Jürgen P. Schulze, Dong Zhang, FACEATLASAR: ATLAS OF FACIAL ACUPUNCTURE POINTS IN AUGMENTED REALITY, Mar 29 2022
 - [8] Tom Bocklisch, Joey Faulkner, Nick Pawlowski, Alan Nichol, Rasa: Open Source Language Understanding and Dialogue Management, Dec 15 2017, pp 1-9.
 - [9] Jacob Devlin Ming-Wei Chang Kenton Lee Kristina Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding Oct 11 2018 , pp.1~16.
 - [10] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, Attention Is All You Need, Jun 12 2017, pp.1-15.
 - [11] crownpku, Rasa NLU for Chinese , Retrieved September 25 , 2022, from https://github.com/crownpku/Rasa_NLU_Chi
 - [12] Alexander Kapitanov, Andrey Makhlyarchuk, Karina Kvanchiani, HaGRID -- HAnd Gesture Recognition Image Dataset, Jun 16 2022. [[CrossRef](#)][[PubMed](#)]
 - [13] Hugging Face (n.d.), BERT model pre-trained for Chinese ,Retrieved December 2 , 2022, from <https://huggingface.co/bert-base-chinese>.