|  |
| --- |
| **國家科學及技術委員會補助專題研究計畫報告** |
|

**以口語表達與手寫表現探討自閉症兒童的行為特徵與學習輔助**

報告類別：成果報告

計畫類別：個別型計畫

計畫編號：MOST 112-2221-E-468-009-

執行期間：112 年 08 月 01 日 至 113 年 12 月 31 日

執行機構及系所：亞洲大學資訊工程學系

計畫主持人：陳良弼

共同主持人：藍先元

|  |
| --- |
| 本研究具有政策應用參考價值： □否 ■是，建議提供機關衛生福利部  (勾選「是」者，請列舉建議可提供施政參考之業務主管機關)  本研究具影響公共利益之重大發現：■否 □是 |

中 華 民 國 112 年 12 月 1 日

中文摘要: 隨著資訊電子化的政策及便利性，醫療單位朝向透過電子病歷記錄病的記錄而 醫療資料經年累月累積形成巨量資料，帶來巨量醫療資料的整合、處理與分析等 研究議題。現今社會忙碌的人也因為網路資訊的發達，可以透過線上諮詢取得專 業醫生的回覆，清楚是否該就醫或獲得其他醫療訊息。而多元的社群媒體平台， 讓人們可以藉此查詢與學習各種知識，也能夠在平台上創作與分享各種內容，造 就社群媒體成為現代生活不可或缺的核心媒介。因此，透過分析醫療資料、社群 媒體資料與問答網站資料，可以產生各式新型醫療服務，不僅能幫助普通民眾了 解自身需求，也能協助醫院與醫生提升病患醫療滿意度，降低醫療資源的浪費。 著眼於此，本計畫以三年為期，探討運用多元資料與深度學習技術於新型醫療服 務之開發。在本年度計畫執行過程中，我們已完成『基於組合特徵的注意力機制 與完全共享的多任務學習之生物醫學專名識別』、『結合知識檢索來進行心理健康 狀況檢測』、『使用字典選擇的高風險文章的詞級模型進行自殺風險評估』、『透過 整合對話歷史和輔導者特徵進行情緒輔導對話的策略預測』共四項研究。本期末 報告茲就本年度所完成的研究成果進行報告。

中文關鍵詞: 輔助醫療決策、資料分析、問答模型、疾病預測模型、情感分析

英文摘要: The convenience of the Internet has made most medical institutions to build electronic health records (EHR). The accumulated EHR over time provides good opportunities for new research. Moreover, people nowadays can easily get online professional medical consultations via Internet in addition to traditional medical treatments. Finally, social media have become an indispensable core media for most people. In this project, we research into medical service applications by analyzing EHR, social media, and Q&A data. The expected results not only assist people on what they need, but also help medical institutions and medical professionals improve the satisfaction of patients and reduce the waste of medical resources. In this progress report, four research results we achieved in this year are presented, including 1) biomedical named entity recognition with the combined feature attention and fully-shared multi-task learning, 2) the detection of mental health conditions by incorporating knowledge retrieval, 3) suicide risk assessment using word-level model with dictionary-based risky posts selection, and 4) predicting the following support strategy during the emotional support dialogue by integrating the dialogue history and supporter features.

英文關鍵詞: Medical Decision Support, Data Analysis, QA Model, Disease Prediction Model, Sentiment Analysis

**科技部補助專題研究計畫成果報告**

**（□期中進度報告 / ■期末報告）**

**以口語表達與手寫表現探討自閉症兒童的行為特徵與學習輔助**

**Identification of Behavioral Characteristics of Autism Children with Their Narrative and Handwriting**

計畫類別：■ 個別型計畫　　□ 整合型計畫

計畫編號：112-2221-E-468-009-

執行期間：2023年08月01日至2024年12月31日

執行機構及系所：亞洲大學資訊工程學系

計畫主持人：陳良弼

計畫參與人員：蔡昀陞、林鎰鋒、林昀昇、嚴翎愷、李昀叡、SYAUKI AULIA THAMRIN

本計畫除繳交成果報告外，另含下列出國報告，共 \_0\_ 份：

□ 執行國際合作與移地研究心得報告

□ 出席國際學術會議心得報告

□ 出國參訪及考察心得報告

中 華 民 國　　112 　年　　12　　月 　 1　　 日

**中文摘要**

隨著資訊電子化的政策及便利性，醫療單位朝向透過電子病歷記錄病人的看診記錄，而醫療資料經年累月累積形成巨量資料，帶來巨量醫療資料的整合、處理與分析等研究議題。現今社會忙碌的人也因為網路資訊的發達，可以透過線上諮詢，取得專業醫生的回覆，清楚是否該就醫或獲得其他醫療訊息。而多元的社群媒體平台，讓人們可以藉此查詢與學習各種知識，也能夠在平台上創作與分享各種內容，造就社群媒體成為現代生活不可或缺的核心媒介。因此，透過分析醫療資料、社群媒體資料與問答網站資料，可以產生各式新型醫療服務，不僅能幫助普通民眾了解自身需求，也能協助醫院與醫生提升病患醫療滿意度，降低醫療資源的浪費。著眼於此，本計畫以三年為期，探討運用多元資料與深度學習技術於新型醫療服務之開發。在本年度計畫執行過程中，我們已完成『基於組合特徵的注意力機制與完全共享的多任務學習之生物醫學專名識別』、『結合知識檢索來進行心理健康狀況檢測』、『使用字典選擇的高風險文章的詞級模型進行自殺風險評估』、『透過整合對話歷史和輔導者特徵進行情緒輔導對話的策略預測』共四項研究。本期末報告茲就本年度所完成的研究成果進行報告。

**Abstract**

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by impairments in social communication and interaction, as well as restricted and repetitive behaviors. ASD affects individuals across various aspects of their lives, including their cognitive, social, and motor skills. Early detection and intervention are crucial for improving outcomes and providing appropriate support to individuals with ASD. With recent advancement regarding computational language models, we aim to utilize machine learning and deep neural network models to analyze children's narratives and handwriting to detect autism tendencies and characteristics. These approaches can contribute to improving diagnostic efficiency and providing comprehensive language evaluations to better understand and support the needs of children with autism. In this progress report, three research results we achieved in this year are presented, including 1) identifying Chinese handwriting characteristics for detecting children with autism, 2) analyzing handwriting characteristics of children with autism by Chinese characters and Mandarin Phonetic Symbols I, 3) using natural language processing to analyze autistic tendencies from children's picture book narratives, with the second research being the successor of the first.

自閉症譜系障礙（ASD）是一種神經發展性障礙，其特徵是社交溝通和互動的缺陷，以及有限且重複的行為。自閉症影響個體生活的各個方面，包括認知、社交和運動技能。早期發現和給予輔助對於自閉症個體的社會化發展至關重要。隨著計算語言模型的進步，我們旨在利用機器學習和深度神經網絡模型來分析兒童的敘事和手寫內容，以檢測自閉症的傾向和特徵。這些方法有助於提高診斷效率，並提供全面的語言評估，以更好地理解和支持自閉症兒童的需求。

在這份進展報告中，我們展示了今年取得的三項研究成果，包括：1) 識別用於檢測自閉症兒童的中文手寫特徵，2) 通過中文字符和注音符號分析自閉症兒童的手寫特徵，3) 利用自然語言處理技術從兒童的繪本敘事中分析自閉症傾向。其中，第二項研究是第一項研究的延續。

**一、前言**

自閉症譜系障礙（Autism Spectrum Disorder, ASD）是一種複雜的神經發育障礙，主要特徵為在社交互動和溝通方面的挑戰，並伴隨著限制性、重複性的行為模式、興趣或活動 [2-1]。這是一種終身的障礙，通常從幼年就開始影響個體在各個功能領域的表現。「譜系」一詞指的是自閉症譜系障礙中症狀、能力和特徵的廣泛變化，這些特徵在不同自閉症患者之間差異很大。

根據自閉症和發展障礙監測（Autism and Developmental Disabilities Monitoring, ADDM）報告中顯示，2020年美國8歲兒童中每千人有27.6例被診斷為自閉症，大約每36名兒童中就有一名被診斷為自閉症。並且此研究顯示自閉症患者性別差異顯著，男性的自閉症發生率是女性的約3.8倍（Maenner等，2023）。2023年，台灣最新的自閉症人數為19,078人，其中74.6%被認定為輕度自閉症。大多數輕度自閉症兒童通常在進入小學後才被發現並診斷。這些自閉症兒童可能會面臨各種學習和社交挑戰，並需要個別的教育支持。因此，幫助學校教師發現具有自閉症特徵的兒童，可能會加快提供適當教育資源的進程。

診斷自閉症的過程漫長且耗時。其中一個原因是因為自閉症的定義主要依賴行為，而非基因上的特徵或物理生物標記（Schaaf等，2020）。自閉症的診斷過程通常涉及多方面的評估。其主要分為兩個階段：篩查和綜合診斷評估。典型的自閉症診斷過程包括多學科的合作，並收集來自各種來源的信息，包括但不限於：(1) 與父母或照護者的訪談，(2) 幼兒園或學校教師的教育日誌或行為觀察，(3) 標準化的行為評估和體格檢查（Goldstein & Ozonoff, 2018）。整個診斷過程通常需要數月的時間，並伴隨著多次的醫院訪問，這對於許多家庭而言可能是巨大的時間和資源負擔。

自閉症個體的一個顯著特徵是他們在不同的社交情境中，持續面臨社交溝通和互動的挑戰（美國精神醫學協會，2022）。在不同語言中，研究一再表明，自閉症個體在理解和社交性地使用語言方面存在缺陷（Engberg-Pedersen & Christensen, 2016；Sah & Torng, 2015；Mäkinen等，2014）。自閉症兒童講故事時的敘事能力是在語言使用方面最常見的問題（Baixauli, Colomer, Roselló, & Miranda, 2016）。研究發現自閉症兒童的敘事往往缺乏連貫性和因果聯繫，並可能包含無關或不符合的內容（Diehl, Bennetto, & Young, 2006）。此外，與同齡典型發展的兒童相比，自閉症兒童在講故事時的語句數量較少，詞彙多樣性也較低（Capps, Losh, & Thurber, 2000）。自閉症兒童無法描述角色的思想和情感，其原因可能是他們無法理解角色行為背後的動機。綜合以上研究與自閉症兒童的敘事特徵，我們可能可以通過兒童的敘事來識別具有自閉症特徵的兒童。

另一方面，自閉症兒童在整合性動作上可能也會遇到困難，進而對他們的日常生活帶來挫折。其中文字書寫為代表性的動作之一。中文字符與英文字母之間的差異涉及多個方面，其中一個顯著的區別在於它們的字形結構 [19]。英文字母的結構相對簡單，通常由少量的曲線和直線組成；相比之下，中文字符則包含豐富的筆劃和複雜的結構，各種子結構和組織模式，深刻表現字符的含義和語境。另一方面，英文字母中常見的曲線和圓形在中文字符中相對較少見，中文字符的主要元素是直線和橫筆劃。這種區別也會影響手寫過程中的運動計劃和協調，對學習者，尤其是自閉症個體來說，可能構成挑戰。因此，研究自閉症兒童在中文書寫中的具體表現，是一項具有挑戰性但意義深遠的課題。

本計畫旨在發現自閉症手寫和敘事中的潛在特徵，分別反映出前述的兩項發現。我們完成了名為「[名稱]」的研究。方法和結論將在後續部分中介紹。

**國內外相關研究**

1. **ADOS-2**

ADOS-2是一種標準化評估工具，用於幫助診斷從12個月到成年的自閉症譜系障礙（Lord等，2012）。在可靠性方面，ADOS-2的評分者間信度(inter-rater reliability)較高，再測信度(test-retest reliability)介於.68到.92之間（李宜融，2015）。ADOS-2由半結構化的活動（例如遊戲和訪談）組成，為評估者提供觀察與ASD診斷相關行為的機會。根據個體的年齡和溝通能力，評估者可以從五個不同模組中選擇合適的模組。一次的評估時長大約為40至60分鐘。

在[第三篇論文]中，我們使用了模組3中的故事講述活動來收集兒童的敘事。兒童需根據指定無字繪本《Tuesday》（Wiesner，1991）講述其故事。實驗的目的是錄製兒童的敘事過程，並使用「人類轉錄分析編碼」（Codes for Human Analysis of Transcripts, CHAT）（MacWhinney & Snow, 1990）轉錄故事為文字。

這項任務評估兒童理解、運用所提供的視覺線索來講述繪本故事的能力。也評估了兒童理解故事順序的能力。

**2、Deep Learning for the Automatic Detection of ASD**

目前已經有許多先前的研究應用深度學習來檢測自閉症，並使用了不同類型的數據。例如，Ahmed等人 [15] 使用了眼動追蹤圖像，而Zhou等人 [16] 則使用了語音頻譜圖。大多數研究集中於利用自閉症的腦成像數據，包括Heinsfeld等人 [17]、Sewani & Kashef [18] 以及Kong等人 [19]，他們的研究都取得了高的準確率。

在手寫數據方面，Hendr等人 [20] 是第一個應用深度學習來檢測自閉症的研究者。Hendr等人收集了104名參與者的數據，其中51名為ASD患者，53名為非ASD患者。每位參與者完成了18項手寫任務，其中12項為文字書寫，6項為圖形繪製。經過數據處理後，這些手寫數據被輸入到深度學習模型中進行分類，最佳準確率達到90.48%。

**3、Handwriting Characteristics of ASD Children**

書寫技能的評估是同時基於書寫成品和書寫過程兩者。其中書寫成品的評估標準包括字母形狀、大小、傾斜度、間距和線條的直線度，這些標準是基於[21], [22]開發的書寫分析量表所達成的共識。

Fuentes等人 [9] 結合了明尼蘇達書寫評估 [14]，發現自閉症兒童在測試的可讀性部分得分顯著低於同齡典型發展（TD）兒童。Beversdorf等人 [23] 評估了自閉症成人與非自閉症成人的書寫樣本，發現自閉症組寫的字母顯著大於控制組。Johnson等人發現，在自閉症兒童的書寫中有較差的空間排列。此外，自閉症兒童的書寫動作表現出顯著的不穩定性 [24], [25]。

然而，這些任務是專門設計來評估書寫技能的，未必能反映兒童在學校中遇到的真實情境。據我們所知，尚無研究利用來自日常學校環境的兒童書寫數據。通過利用此環境中的書寫數據，我們不僅能更方便地獲取數據(使用現有的書寫練習本，所以參與者無需進行額外的書寫實驗)，還可以觀察與日常環境更密切相關的書寫特徵。這種方法確保了書寫樣本能夠反映兒童在常規學校環境中自然且真實的書寫行為，從而提供更貼近現實情境的書寫特徵表現。

**4、****Computational models for analyzing narration**

先前針對兒童敘事能力的研究主要使用手動分析來探討不同發展階段和不同語言能力的群體中的語言特徵。只有少數研究使用自然語言處理(Natural Language Processing, NLP)工具或神經網路來調查自閉症個體與典型發展（TD）同齡人之間的敘事表現差異。

為了量化描述自閉症個體的敘事表現，Chojnicka和Wawer（2020）採用了基於語言類別模型（Linguistic Category Model）的情感與語言抽象分析。實驗包括50名講波蘭語的兒童（25名自閉症個體和25名TD對照組，年齡範圍7至25歲）。語言樣本來自ADOS-2的兩項標準化任務：「根據書本講故事」和「描述圖片」。結果顯示，情感和語言抽象分析是有價值的工具、自閉症個體的語言抽象能力水平低於TD個體，並且在表達情感方面表現出困難。

在Chojnicka和Wawer的研究中，他們收集了50名講波蘭語個體的敘事樣本（25名自閉症參與者和25名TD對照組，年齡範圍7至25歲），目的是識別自閉症個體所產生的敘事。研究使用了兩個文本編碼器：語言模型嵌入（ELMo）和通用句子編碼器（USE），以及三種分類算法：XGBoost、支持向量機和密集神經網絡層（Wawer & Chojnicka，2022）。實驗表明，使用深度神經網路模型的分析相較人工評分具有更高的敏感度、特異性、陽性預測值和陰性預測值。然而，這些值低於目前的兩個標準化工具：ADOS-2和社交溝通問卷（SCQ）（Rutter, Bailey, & Lord, 2003）。其中SCQ包含40個問題，由熟悉受測者的父母、照護者或教師回答。這些問題涵蓋了社交溝通的各個方面，例如社交互動、語言與非語言溝通、以及與自閉症相關的限制性和重複性行為。SCQ中的每個項目根據是否存在與ASD相關的特定行為來評分，總分可顯示個體可能患有自閉症的可能性，分數越高表明該情況的可能性越大。

**二、研究方法、進行步驟及執行進度報告**

我們擬探討多元資料與深度學習技術於新型醫療服務之開發，並著重於三大主軸：『醫療資料』、『社群媒體』、『問答網站』。第三年成果報告如下所示:

**成果報告：paper 1**

**研究目的**

我們的目標是識別能夠幫助檢測自閉症譜系障礙（ASD）的獨特或共同的書寫特徵。我們使用了多種機器學習模型來對自閉症兒童和典型發展（TD）兒童的書寫進行分類。在分類過程中，模型會嘗試識別書寫特徵作為依據，區分書寫者是自閉症兒童或典型發展兒童。接著，我們研究自閉症兒童在書寫中文和英文時的書寫特徵，以確定這兩種書寫系統之間是否存在相似性或差異性。

然而，據我們所知，目前尚無同時具有自閉症譜系障礙、兒童及繁體中文三個標籤的書寫數據集。因此數據的收集是我們首先要解決的問題。我們與當地的小學合作，收集學生在課堂上使用的書寫練習簿。最終機器學習模型的目標是一個二元分類任務：確定一個漢字書寫圖像是由自閉症兒童還是典型發展兒童書寫的。

**研究方法**

我們收集了參與者以往和目前使用的書寫練習簿。書寫練習簿如所示，這是台灣小學生用來練習書寫漢字的練習簿(圖4)。為了確保模型不受外部因素影響，我們引入了一個「髒亂」標籤來區分「乾淨」與「髒亂」的書寫。0表示乾淨的書寫，沒有額外標記；相反，1表示有糾正字跡（如紅筆標記）或鄰格超出字格範圍的書寫。

表1顯示了參與者的統計數據。本研究包括5名自閉症兒童和17名典型發展兒童，共22名兒童參與。自閉症兒童的平均年齡為11.1歲，而典型發展兒童的平均年齡為8.67歲。性別比方面，典型發展兒童相對均衡，有8名男孩和9名女孩；然而所有自閉症兒童均為男孩。我們收集了共39本書寫練習簿，32本來自典型發展兒童、7本來自閉症兒童。

由於本研究中自閉症和典型發展兒童的數據不平衡問題，我們採用了下採樣(down sampling)技術來處理這一問題。

我們建立了四個資料集：全資料集（Dataset1）、去除「髒亂」書寫的資料集（Dataset2）、平衡資料集（Dataset3，為採用下採樣的資料集）和去除「髒亂」書寫的平衡資料集（Dataset4）。表2展示了這四個數據集的統計數據和屬性。

我們採用了支持向量機（SVM）模型【26】，以及兩個神經網路模型，即LeNet【27】和ResNet-18【28】來進行分類。

SVM是一種廣泛使用的機器學習演算法，常用於分類和迴歸任務。SVM的主要概念是在高維特徵空間中找到能夠將不同類別數據點分開的最佳超平面。SVM提供了一種成熟且可解釋的二元分類方法，並能有效處理小數據集，從而實現良好的泛用性能。

LeNet是一種卷積神經網路（CNN）【29】，是深度學習領域的先驅模型之一，最初設計用於識別手寫數字。LeNet模型包含多層結構，包括卷積層、池化層和全連接層。輸入圖像依次經過卷積層提取相關特徵，經過池化層降低空間維度，最終通過全連接層進行分類。由於LeNet在手寫字符識別方面的成功，我們選擇其作為模型之一。

ResNet-18是另一種深度卷積神經網路，在各種電腦視覺任務中取得了顯著成功。ResNet-18廣泛應用於圖像分類、物體檢測和語義分割等任務。其深層結構和有效的殘差連接使其能夠捕捉細節，並在多個基準數據集中達到領先表現。在本研究中，我們選擇ResNet-18作為分析和分類手寫圖像的模型，因其深度架構能有效處理手寫任務的複雜性。

轉移學習(transfer learning)是一種強大的技術，能夠將知識從一個任務轉移到另一個任務，特別適用於數據有限或需要大量計算資源的任務。圖8所示，我們通過微調預訓練於大規模數據集（如ImageNet【30】）的ResNet-18模型來有效提取手寫數據集相關的特徵。這種方法不僅幫助我們克服數據集小的限制，還能從大型預訓練數據集中繼承豐富的通用知識。

類別激活映射（Class Activation Map, CAM）【31】是一種常用於可視化和解釋深度學習模型（尤其是CNN）的技術。CAM的主要目的是識別輸入圖像中對CNN模型分類決策貢獻最大的區域。通過不同顏色顯示對模型預測有影響的重要區域，CAM為我們提供了更深入的見解，幫助我們理解ASD兒童書寫特徵的區別。

**實驗結果**

我們的SVM模型的表現如表三所示。從“SVM-Dataset1”與“SVM-Dataset2”比較中可以看出移除不乾淨的字符導致F1分數下降。但在另一組中，移除不乾淨的字符卻導致F1分數上升。從“SVM-Dataset1”與“SVM-Dataset3”的比較與“SVM-Dataset2”與“SVM-Dataset4”的比較中可以看出，使用下採樣導致F1分數下降。

LeNet模型的表現如表四所示。與前面的分析類似，我們從“髒字符”和“平衡”的角度分析模型的表現。我們觀察到，移除不乾淨的字符後F1分數提高，這表明沒有修改痕跡或其他額外書寫痕跡的書寫能帶來更準確的模型預測。使用下採樣平衡過後則使F1分數降低。

我們的ResNet-18模型的表現如表五所示。我們觀察到，當移除不乾淨的字符時，F1分數略有下降。但在“平衡”方面，不論是否移除不乾淨的字符，平衡數據集後的性能顯著優於未平衡數據集的性能。這是可能因為平衡數據集後，召回率增加，表明模型更可能預測出由ASD兒童書寫的字跡。這與我們的預期一致，通過平衡數據集來改善資料及不平衡的問題。

接下來，我們比較了不同模型在各數據集上的表現，如表六所示。在三個模型中，LeNet在所有數據集上的表現最差，可能原因是其架構較為簡單。在Dataset1和Dataset2上，SVM分別比ResNet-18高出5%和4%。然而在Dataset3和Dataset4上，ResNet-18顯著優於SVM模型，分別超過22%和10%。這可以歸因於ResNet-18利用了預訓練權重，能夠從大型數據集中借鑒知識。ResNet-18的架構也適合捕捉複雜的書寫特徵，從而實現更準確的預測。

CAM（類激活映射）的結果如圖9、10、11和12所示。在圖9中，我們觀察到，當模型預測字跡由ASD兒童書寫時，它往往依賴於特定的已識別特徵（如圖中用紅圈標記的部分）。當模型預測字跡由TD兒童書寫時，它則傾向於考慮整體字跡，而不是集中在某個部分。除此之外，我們識別出兩個ASD兒童書寫的關鍵特徵。第一，ASD兒童在寫方向突然變化的筆畫時往往會遇到困難。第二，ASD兒童的線條通常比較無法對齊。

實驗結果顯示，我們的模型能夠有效區分ASD兒童與TD兒童的字跡。我們的模型達到的最佳F1分數為93.6%，這表明其在識別ASD兒童獨特書寫特徵方面具有很高的準確性。

**成果報告：paper 2**

**研究目的**

Writing is a crucial aspect of language learning for children, and Chinese character writing places higher demands on hand-eye coordination and motor control.

Our first objective is to incorporate phonetic notation data into this study. We aim to investigate whether the inclusion of phonetic notation data improves the model's performance in classifying autistic and TD children, as in the first study, we did not utilize phonetic notation data.

Our second objective is to design a neatness label to distinguish whether a Chinese character is written in a neat manner. It is a binary label (yes, no) annotated through manual labeling. The purpose of this is to track the neatness of handwriting for both autistic and TD children. By using only neatly written Chinese characters, we plan to train the classification model for the writing from autistic and TD children. This constitutes a more challenging task, as distinguishing the handwriting characteristics between autistic and TD children becomes even more intricate when all Chinese characters are written neatly. The purpose of this approach is to further assess the model's ability to differentiate the handwriting styles of autistic children from TD children when only neatly written Chinese characters are considered.

**研究方法**

This study used the same dataset collected and introduced in Yen et al. [20].

The Chinese character-only dataset comprises 17,950 words, with 14,173 from TD children and 3,777 from autistic children. We considered neatness of writing Chinese characters by defining the neatness criteria based on relevant literature [34, 35] and discussions with experienced elementary school teachers.

The neatness criteria were divided into two levels: stroke and component, and three factors: position, size, and correctness. The division into stroke and component levels allows for a more nuanced and comprehensive assessment of neatness. The neatness is labeled as “1” if the individual word satisfies five or more aspects, and “0” if it satisfies four or less aspects. In conclusion, the author labeled all the pictures in the Chinese character-only dataset with 14,840 having a neatness label of 1 and 3,110 having a neatness label of 0. Detailed information is presented in Table 1. The value of the Chi-square test is 8824.91, with a *p*-value < 0.01, which shows TD children were more likely to produce neat Chinese characters than autistic children.

In pursuit of our first objective of assessing the potential contribution of phonetic notations to the task at hand, we evaluated the impact of incorporating phonetic notations in the classification of the handwritings. The resultant dataset, denoted as the Phonetic notation-only dataset, encompassed a total of 18,833 images. Within this dataset, 14,943 instances pertained to TD children, while 3,890 instances were from autistic children.

In total, there are 3 datasets. The first is the Chinese character-only dataset; the second is the Phonetic notation-only dataset; the third dataset is the Chinese character + Phonetic notation dataset. Table 2 provides detailed number of images in each of the three datasets. The reduced number of instances in the Chinese character + Phonetic notation dataset came from the fact that there are data with only Chinese characters (Figure 5(a)), or only phonetic notations (Figure 5(b)).

Our flowchart of building classification models is divided into two types: a 5-fold flowchart, as shown in Figure 6, and the CAM flowchart, depicted in Figure 7. The 5-fold flowchart signifies our utilization of 5-fold cross-validation during model training. We used the average of the results from the 5 folds for evaluation. On the other hand, the CAM flowchart employed the results of applying CAM for identifying the handwriting characteristics. This required an individual model rather than an average of 5 models.

As seen in Table 1, our data was imbalanced. To address this issue, we applied oversampling and undersampling techniques to the data in the training set.

In terms of models, we employed support vector machine (SVM) [39], decision tree (DT) [40], K-nearest neighbor (KNN) [41], and logistic regression (LR) [42]—commonly used models in machine learning that are applicable to classification tasks. Additionally, we also utilized the ResNet-18 [38] model, mentioned in the first study.

Class Activation Map (CAM) [44] serves as a crucial visualization tool that facilitates a deeper understanding of a model's focal points. This technique involves establishing a connection with the Global Average Pooling (GAP) layer after the final convolutional layer. Following this connection, it captures the weights associated with the GAP layer output and linearly combines them with the corresponding feature map to generate the results. The conventional approach outlined above mandates the utilization of the GAP layer, imposing constraints on the overall flexibility of the network architecture. To overcome this limitation, Grad-CAM [23] introduces an innovative solution by incorporating the partial differential of the feature map in the relevant category to supplant the weight output derived from the GAP layer. This modification enables Grad-CAM to be applied across a broader spectrum of CNN architectures, and therefore offers enhanced adaptability. In the context of this study, we adopted the Grad-CAM approach, aligning with the methodology employed by the first study. This ensured consistency and enabled us to leverage the proven effectiveness of Grad-CAM in visualizing and interpreting the focus areas of our model.

**實驗結果**

In the neatness classification experiments, we utilized SVM, DT, KNN, LR, and ResNet-18 models, each with three methods to deal with the data imbalance problem: undersampling, oversampling, and X which means neither undersampling nor oversampling techniques were applied. All results were presented in Table 3, and the best performance of each model was summarized in Table 4.

**成果報告：paper 3**

**研究目的**

**研究方法**

**實驗結果**

**三、成果自評**

本計畫為三年期計畫之最後一年，在本年度的計畫執行過程中，我們已經順利完成生物醫學專名識別、心理健康狀況檢測、自殺風險評估、情緒輔導對話的策略預測共四項研究。成果包含相關研究論文4篇，其中三篇分別被國際期刊BMC bioinformatics、Journal of Intelligent Information Systems、Multimedia Tools and Applications接受。

三年的時間下來，我們深刻體會到醫療和科技技術的結合並不是一件簡單的事，

當中包括資料的處理和最後結果的解釋，都需要不同背景的專業知識和一定經驗的累積，我們也相信還有許多應用尚未被發展出來，但這三年的時間，我們投入了許多時間和心力在分析醫療資料、社群媒體資料和問答網站資料，並運用深度學習技術在各個應用上，例如，出加護病房預測、情緒原因分析、自殺風險評估等等，當中許多成果也被國際期刊所接受，這已經達到了我們最一開始的預期成果，也代表了我們的成果在這方面做出了一定的貢獻。

**已接受、發表之論文**

1. Zhang, Z., & Chen, A. L. (2022). Biomedical named entity recognition with the combined feature attention and fully-shared multi-task learning. BMC bioinformatics, 23(1), 1-21.

2. Lin, Y. S., Tai, L. K., & Chen, A. L. (2023). The detection of mental health conditions by incorporating external knowledge. Journal of Intelligent Information Systems, 1-22.

3. Tsai, Y. S., & Chen, A. L. (2023). Suicide risk assessment using word-level model with dictionary-based risky posts selection. Multimedia Tools and Applications, 1-20.

**碩士畢業論文**

1. Zhi Yu Zhang, “Biomedical Named Entity Recognition with the Combined Feature Attention and Fully-Shared Multi-Task Learning,” National Tsing Hua University, 2022.

2. Yun Sheng Lin, “The Detection of Mental Health Conditions by Incorporating Knowledge Retrieval,” National Tsing Hua University, 2022.

3. Yun Sheng Tsai, “Suicide Risk Assessment using Word-Level Model with Dictionary-Based Risky Posts Selection,” National Tsing Hua University, 2022.

4. Yi Feng Lin, “Predicting the Following Support Strategy during the Emotional Support Dialogue by Integrating the Dialogue History and Supporter Features,” National Tsing Hua University, 2022.

**參考文獻**

[1] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv*, p. preprint arXiv:1810.04805.

[2] Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, (p. 32).

[3] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., & ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv*, p. preprint arXiv:1907.11692.

[4] Lee, J., Yoon, W., Kim, S., Kim, D. K., So, C. H., & Kang, J. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, pp. 36(4), 1234-1240.

[5] Tian, Y., Shen, W., Song, Y., & Xia, F. (2020). Improving biomedical named entity recognition with syntactic information. *BMC bioinformatics*, pp. 21(1), 1-17.

[6] Miller, A., Fisch, A., Dodge, J., Karimi, A. H., Bordes, A., & Weston, J. (2016). Key-value memory networks for directly reading documents. *In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, (pp. 1400-1409).

[7] Tian, Y., Song, Y., Ao, X., Xia, F., Quan, X., Zhang, T., & Wang, Y. (2020). Joint Chinese word segmentation and part-of-speech tagging via two-way attentions of auto-analyzed knowledge. *In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, (pp. 8286-8296).

[8] Crichton, G., Pyysalo, S., Chiu, B., & Korhonen, A. (2017). A neural network multi-task learning approach to biomedical named entity recognition. *BMC bioinformatics*, pp. 18(1), 1-14.

[9] Chai, Z., Jin, H., Shi, S., Zhan, S., Zhuo, L., & Yang, Y. (2022). Hierarchical shared transfer learning for biomedical named entity recognition. *BMC bioinformatics*, pp. 23(1), 1-14.

[10] Huang, K., Huang, D., Liu, Z., & Mo, F. (2020, November). A Joint Multiple Criteria Model in Transfer Learning for Cross-domain Chinese Word Segmentation. *In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, (pp. 3873-3882).

[11] Arman Cohan, Bart Desmet, Andrew Yates, Luca Soldaini, Sean MacAvaney, and Nazli Goharian. SMHD: a Large-Scale Resource for Exploring Online Language Usage for Multiple Mental Health Conditions. in *Proceedings of the 27th International Conference on Computational Linguistics*. 2018.

[12] Zheng Ping Jiang, Sarah Ita Levitan, Jonathan Zomick, and Julia Hirschberg. Detection of mental health from reddit via deep contextualized representations. in *Proceedings of the 11th International Workshop on Health Text Mining and Information Analysis*. 2020.

[13] Ankit Murarka, Balaji Radhakrishnan, and Sushma Ravichandran. Classification of mental illnesses on social media using RoBERTa. in *Proceedings of the 12th International Workshop on Health Text Mining and Information Analysis*. 2021.

[14] Tao Gui, Liang Zhu, Qi Zhang, Minlong Peng, Xu Zhou, Keyu Ding, and Zhigang Chen. Cooperative multimodal approach to depression detection in twitter. in *Proceedings of the AAAI conference on artificial intelligence*. 2019.

[15] Guangyao Shen, Jia Jia, Liqiang Nie, Fuli Feng, Cunjun Zhang, Tianrui Hu, Tat-Seng Chua, and Wenwu Zhu. Depression detection via harvesting social media: A multimodal dictionary learning solution. in *IJCAI*. 2017.

[16] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. Predicting depression via social media. in *Seventh international AAAI conference on weblogs and social media*. 2013.

[17] Glen Coppersmith, Mark Dredze, and Craig Harman. Quantifying mental health signals in Twitter. in *Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality*. 2014.

[18] Andrew G Reece and Christopher M Danforth, Instagram photos reveal predictive markers of depression*.* EPJ Data Science, 2017. **6**(1): p. 15.

[19] Marjan Ghazvininejad, Chris Brockett, Ming-Wei Chang, Bill Dolan, Jianfeng Gao, Wen-tau Yih, and Michel Galley. A knowledge-grounded neural conversation model. in *Proceedings of the AAAI Conference on Artificial Intelligence*. 2018.

[20] Dongfang Li, Baotian Hu, Qingcai Chen, Weihua Peng, and Anqi Wang. Towards medical machine reading comprehension with structural knowledge and plain text. in *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*. 2020.

[21] Vladimir Karpukhin, Barlas Oguz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. Dense Passage Retrieval for Open-Domain Question Answering. in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2020.

[22]Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, and Tim Rocktäschel, Retrieval-augmented generation for knowledge-intensive nlp tasks*.* Advances in Neural Information Processing Systems, 2020. **33**: p. 9459-9474.

[23] Shing, H.-C., Nair, S., Zirikly, A., Friedenberg, M., Daumé III, H., & Resnik, P. (2018). Expert, crowdsourced, and machine assessment of suicide risk via online postings. Proceedings of the fifth workshop on computational linguistics and clinical psychology: from keyboard to clinic.

[24] Matero, M., Idnani, A., Son, Y., Giorgi, S., Vu, H., Zamani, M., Limbachiya, P., Guntuku, S. C., & Schwartz, H. A. (2019). Suicide risk assessment with multi-level dual-context language and BERT. Proceedings of the sixth workshop on computational linguistics and clinical psychology.

[25] Gaur, M., Alambo, A., Sain, J. P., Kursuncu, U., Thirunarayan, K., Kavuluru, R., Sheth, A., Welton, R., & Pathak, J. (2019). Knowledge-aware assessment of severity of suicide risk for early intervention. The world wide web conference.

[26] Sawhney, R., Joshi, H., Gandhi, S., & Shah, R. R. (2021). Towards ordinal suicide ideation detection on social media. Proceedings of the 14th ACM International Conference on Web Search and Data Mining.

[27] Sawhney, R., Neerkaje, A. T., & Gaur, M. (2022). A Risk-Averse Mechanism for Suicidality Assessment on Social Media. *Association for Computational Linguistics 2022 (ACL 2022)*.

[28] Zhou, H., et al. Emotional chatting machine: Emotional conversation generation with internal and external memory. in *Proceedings of the AAAI Conference on Artificial Intelligence*. 2018.

[29] Zhou, X. and W.Y. Wang. MojiTalk: Generating Emotional Responses at Scale. in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2018.

[30] Ghosal, D., et al. DialogueGCN: A Graph Convolutional Neural Network for Emotion Recognition in Conversation. in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 2019.

[31] Sharma, A., et al. A Computational Approach to Understanding Empathy Expressed in Text-Based Mental Health Support. in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2020.

[32] Majumder, N., et al. MIME: MIMicking Emotions for Empathetic Response Generation. in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2020.

[33] O'Leary, K., et al. “Suddenly, we got to become therapists for each other” Designing Peer Support Chats for Mental Health. in *Proceedings of the 2018*

[34] Hill, C.E., K.M. O'Brien, and K. O'Brien, Helping skills: Facilitating exploration, insight, and action. 1999: American Psychological Association Washington, DC.

[35] Sun, H., et al. PsyQA: A Chinese Dataset for Generating Long Counseling Text for Mental Health Support. in *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*. 2021.

[36] Liu, S., et al. Towards Emotional Support Dialog Systems. in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 2021.

[37] Peng, W., et al., Control Globally, Understand Locally: A Global-to-Local Hierarchical Graph Network for Emotional Support Conversation*.* arXiv preprint arXiv:2204.12749, 2022.

[38] Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., & McClosky, D. (2014, June). The Stanford CoreNLP natural language processing toolkit. *In Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, (pp. 55-60).

[39] World Health Organization, The WHO special initiative for mental health (2019-2023): universal health coverage for mental health. 2019, World Health Organization.

[40] World Health Organization, Comprehensive mental health action plan 2013–2030. 2021, World Health Organization,, .

[41] Z. Steel, C. Marnane, C. Iranpour, T. Chey, J. W. Jackson, V. Patel, and D. Silove, The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013*.* Int J Epidemiol, 2014. **43**(2): p. 476-93.

[42] V. Patel, S. Saxena, C. Lund, G. Thornicroft, F. Baingana, P. Bolton, D. Chisholm, P. Y. Collins, J. L. Cooper, J. Eaton, H. Herrman, M. M. Herzallah, Y. Huang, M. J. D. Jordans, A. Kleinman, M. E. Medina-Mora, E. Morgan, U. Niaz, O. Omigbodun, M. Prince, A. Rahman, B. Saraceno, B. K. Sarkar, M. De Silva, I. Singh, D. J. Stein, C. Sunkel, and J. UnUtzer, The Lancet Commission on global mental health and sustainable development*.* Lancet, 2018. **392**(10157): p. 1553-1598.

[43] V. Bird, P. Premkumar, T. Kendall, C. Whittington, J. Mitchell, and E. Kuipers, Early intervention services, cognitive-behavioural therapy and family intervention in early psychosis: systematic review*.* Br J Psychiatry, 2010. **197**(5): p. 350-6.

[44] J. Treasure and G. Russell, The case for early intervention in anorexia nervosa: theoretical exploration of maintaining factors*.* Br J Psychiatry, 2011. **199**(1): p. 5-7.

[45] Glen Coppersmith, Mark Dredze, Craig Harman, and Kristy Hollingshead. From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses. in *Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality*. 2015.

[46] Sean Macavaney, Anjali Mittu, Glen Coppersmith, Jeff Leintz, and Philip Resnik. Community-level research on suicidality prediction in a secure environment: Overview of the CLPsych 2021 shared task. in *Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access*. 2021.

[47] Adrian Benton, Margaret Mitchell, and Dirk Hovy. Multitask Learning for Mental Health Conditions with Limited Social Media Data. in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. 2017.

[48] James N Butcher, John R Graham, Yossef S Ben-Porath, Auke Tellegen, WG Dahlstrom, and Beverly Kaemmer, MMPI-2: Manual for administration, scoring, and interpretation (Rev. ed.)*.* Minneapolis, MN: University of Minnesota, 2001.

[49] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders. 5th ed. 2013, Washington, DC.

[50] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, Attention Is All You Need*.* Advances in Neural Information Processing Systems 30 (Nips 2017), 2017. **30**.

[51] Ivan Sekulic and Michael Strube, Adapting Deep Learning Methods for Mental Health Prediction on Social Media*.* W-NUT 2019, 2019. **14**(162.2): p. 322.

[52]李明濱. (2020). *109年自殺防治年報*. <https://www.tsos.org.tw/media/4591#doc-tabs-detail>

[53] Cao, L., Zhang, H., Feng, L., Wei, Z., Wang, X., Li, N., & He, X. (2019). Latent suicide risk detection on microblog via suicide-oriented word embeddings and layered attention. *arXiv preprint arXiv:1910.12038*.

[54] Olfson, M., Building the mental health workforce capacity needed to treat adults with serious mental illnesses. Health Affairs, 2016. 35(6): p. 983-990.

## 111年度專題研究計畫成果彙整表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **計畫主持人：**陳良弼 | | | **計畫編號：**109-2221-E-468-014-MY3 | | |
| **計畫名稱：**基於多元資料及深度學習技術之醫療服務應用 | | | | | |
| 成果項目 | | | 量化 | 單位 | 質化  （說明：各成果項目請附佐證資料或細項說明，如期刊名稱、年份、卷期、起訖頁數、證號...等） |
| 國內 | 學術性論文 | 期刊論文 | 0 | 篇 |  |
| 研討會論文 | 0 |  |
| 專書 | 4 | 本 | 1. Zhi Yu Zhang, “Biomedical Named Entity Recognition with the Combined Feature Attention and Fully-Shared Multi-Task Learning,” National Tsing Hua University, 2022. 2. Yun Sheng Lin, “The Detection of Mental Health Conditions by Incorporating Knowledge Retrieval,” National Tsing Hua University, 2022. 3. Yun Sheng Tsai, “Suicide Risk Assessment using Word-Level Model with Dictionary-Based Risky Posts Selection,” National Tsing Hua University, 2022. 4. Yi Feng Lin, “Predicting the Following Support Strategy during the Emotional Support Dialogue by Integrating the Dialogue History and Supporter Features,” National Tsing Hua University, 2022. |
| 專書論文 | 0 | 章 |  |
| 技術報告 | 0 | 篇 |  |
| 其他 | 0 | 篇 |  |
| 國外 | 學術性論文 | 期刊論文 | 3 | 篇 | 1. Zhang, Z., & Chen, A. L. (2022). Biomedical named entity recognition with the combined feature attention and fully-shared multi-task learning. BMC bioinformatics, 23(1), 1-21.. 2. Lin, Y. S., Tai, L. K., & Chen, A. L. (2023). The detection of mental health conditions by incorporating external knowledge. Journal of Intelligent Information Systems, 1-22. 3. Tsai, Y. S., & Chen, A. L. (2023). Suicide risk assessment using word-level model with dictionary-based risky posts selection. Multimedia Tools and Applications, 1-20. |
| 研討會論文 | 0 |  |
| 專書 | 0 | 本 |  |
| 專書論文 | 0 | 章 |  |
| 技術報告 | 0 | 篇 |  |
| 其他 | 0 | 篇 |  |
| 參與計畫人力 | 本國籍 | 大專生 | 0 | 人次 |  |
| 碩士生 | 5 | 蔡昀陞、林鎰鋒、林昀昇、嚴翎愷、李昀叡 |
| 博士生 | 0 |  |
| 博士級研究人員 | 0 |  |
| 專任人員 | 0 |  |
| 非本國籍 | 大專生 | 0 |  |
| 碩士生 | 0 |  |
|  |  |  |
| 博士生 | 1 | SYAUKI AULIA THAMRIN |
| 博士級研究人員 | 0 |  |
| 專任人員 | 0 |  |
| 其他成果  （無法以量化表達之成果如辦理學術活動  、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。） | | |  | | |